

Wen Yu
Haibo He
Nian Zhang (Eds.)

LNCS 5553

Advances in Neural Networks – ISNN 2009

6th International Symposium on Neural Networks, ISNN 2009
Wuhan, China, May 2009
Proceedings, Part III

3
Part III

 Springer

Commenced Publication in 1973

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison

Lancaster University, UK

Takeo Kanade

Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler

University of Surrey, Guildford, UK

Jon M. Kleinberg

Cornell University, Ithaca, NY, USA

Alfred Kobsa

University of California, Irvine, CA, USA

Friedemann Mattern

ETH Zurich, Switzerland

John C. Mitchell

Stanford University, CA, USA

Moni Naor

Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz

University of Bern, Switzerland

C. Pandu Rangan

Indian Institute of Technology, Madras, India

Bernhard Steffen

University of Dortmund, Germany

Madhu Sudan

Massachusetts Institute of Technology, MA, USA

Demetri Terzopoulos

University of California, Los Angeles, CA, USA

Doug Tygar

University of California, Berkeley, CA, USA

Gerhard Weikum

Max-Planck Institute of Computer Science, Saarbruecken, Germany

Wen Yu Haibo He Nian Zhang (Eds.)

Advances in Neural Networks – ISNN 2009

6th International Symposium
on Neural Networks, ISNN 2009
Wuhan, China, May 26-29, 2009
Proceedings, Part III

Volume Editors

Wen Yu

Centro de Investigación y de Estudios Avanzados
del Instituto Politécnico Nacional (CINVESTAV-IPN)
Departamento de Control Automático
A.P. 14-740, Av. IPN 2508, 07360 México D.F., Mexico
E-mail: yuw@ctrl.cinvestav.mx

Haibo He

Stevens Institute of Technology
Department of Electrical and Computer Engineering
Castle Point on Hudson, Hoboken, NJ 07030, USA
E-mail: hhe@stevens.edu

Nian Zhang

South Dakota School of Mines & Technology
Department of Electrical and Computer Engineering
501 East St. Joseph Street, Rapid City, SD 57701, USA
E-mail: nian.zhang@sdsmt.edu

Library of Congress Control Number: Applied for

CR Subject Classification (1998): F.1, F.2, D.1, G.2, I.2, C.2, I.4-5, J.1-4

LNCS Sublibrary: SL 1 – Theoretical Computer Science and General Issues

ISSN 0302-9743
ISBN-10 3-642-01512-3 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-01512-0 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

springer.com

© Springer-Verlag Berlin Heidelberg 2009
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12672110 06/3180 5 4 3 2 1 0

Preface

This book and its companion volumes, LNCS vols. 5551, 5552 and 5553, constitute the proceedings of the 6th International Symposium on Neural Networks (ISNN 2009), held during May 26–29, 2009 in Wuhan, China. Over the past few years, ISNN has matured into a well-established premier international symposium on neural networks and related fields, with a successful sequence of ISNN symposia held in Dalian (2004), Chongqing (2005), Chengdu (2006), Nanjing (2007), and Beijing (2008). Following the tradition of the ISNN series, ISNN 2009 provided a high-level international forum for scientists, engineers, and educators to present state-of-the-art research in neural networks and related fields, and also to discuss with international colleagues on the major opportunities and challenges for future neural network research.

Over the past decades, the neural network community has witnessed tremendous efforts and developments in all aspects of neural network research, including theoretical foundations, architectures and network organizations, modeling and simulation, empirical study, as well as a wide range of applications across different domains. The recent developments of science and technology, including neuroscience, computer science, cognitive science, nano-technologies and engineering design, among others, have provided significant new understandings and technological solutions to move the neural network research toward the development of complex, large-scale, and networked brain-like intelligent systems. This long-term goal can only be achieved with the continuous efforts of the community to seriously investigate different issues of the neural networks and related fields. To this end, ISNN 2009 provided a great platform for the community to share their latest research results, discuss critical future research directions, stimulate innovative research ideas, as well as facilitate international multidisciplinary collaborations.

ISNN 2009 received 1235 submissions from about 2459 authors in 29 countries and regions (Australia, Brazil, Canada, China, Democratic People's Republic of Korea, Finland, Germany, Hong Kong, Hungary, India, Islamic Republic of Iran, Japan, Jordan, Macao, Malaysia, Mexico, Norway, Qatar, Republic of Korea, Singapore, Spain, Taiwan, Thailand, Tunisia, UK, USA, Venezuela, Vietnam, and Yemen) across six continents (Asia, Europe, North America, South America, Africa, and Oceania). Based on the rigorous peer reviews by the Program Committee members and the reviewers, 409 high-quality papers were selected for publication in the LNCS proceedings, with an acceptance rate of 33.1%. These papers cover major topics of the theoretical research, empirical study, and applications of neural networks. In addition to the contributed papers, the ISNN 2009 technical program included five plenary speeches by Anthony Kuh (University of Hawaii at Manoa, USA), Jose C. Principe (University of Florida, USA), Leszek Rutkowski (Technical University of Czestochowa, Poland), Fei-Yue Wang (Institute of Automation, Chinese Academy of Sciences, China) and Cheng Wu (Tsinghua University, China). Furthermore, ISNN 2009 also featured five special sessions focusing on emerging topics in neural network research.

As organizers of ISNN 2009, we would like to express our sincere thanks to the Huazhong University of Science and Technology, The Chinese University of Hong Kong, and the National Natural Science Foundation of China for their sponsorship, to the IEEE Wuhan Section, the IEEE Computational Intelligence Society, the International Neural Network Society, the Asia Pacific Neural Network Assembly, and the European Neural Network Society for their technical co-sponsorship, and to the Systems Engineering Society of Hubei Province and the IEEE Hong Kong Joint Chapter on Robotics and Automation and Control Systems for their logistic support.

We would also like to sincerely thank the General Chair and General Co-chairs for their overall organization of the symposium, members of the Advisory Committee and Steering Committee for their guidance in every aspect of the entire conference, and the members of the Organizing Committee, Special Sessions Committee, Publication Committee, Publicity Committee, Finance Committee, Registration Committee, and Local Arrangements Committee for all their great effort and time in organizing such an event. We would also like to take this opportunity to express our deepest gratitude to the members of the International Program Committee and all reviewers for their professional review of the papers; their expertise guaranteed the high quality of technical program of ISNN 2009!

Furthermore, we would also like to thank Springer for publishing the proceedings in the prestigious series of *Lecture Notes in Computer Science*. Moreover, we would like to express our heartfelt appreciations to the plenary and panel speakers for their vision and discussion of the latest research developments in the field as well as critical future research directions, opportunities, and challenges.

Finally, we would like to thank all the speakers, authors, and participants for their great contribution and support that made ISNN 2009 a great success.

May 2009

Wen Yu
Haibo He
Nian Zhang

Organization

General Chair

Shuzi Yang, China

General Co-chairs

Youlun Xiong, China

Yongchuan Zhang, China

Advisory Committee Chairs

Shoujue Wang, China

Paul J. Werbos, USA

Advisory Committee Members

Shun-ichi Amari, Japan

Zheng Bao, China

Tianyou Chai, China

Guanrong Chen, China

Shijie Cheng, China

Ruwei Dai, China

Jay Farrell, USA

Chunbo Feng, China

Russell Eberhart, USA

David Fogel, USA

Walter J. Freeman, USA

Kunihiko Fukushima, Japan

Marco Gilli, Italy

Aike Guo, China

Xingui He, China

Zhenya He, China

Petros Loannou, USA

Janusz Kacprzyk, Poland

Nikola Kasabov, New Zealand

Okyay Kaynak, Turkey

Frank L. Lewis, USA

Deyi Li, China

Yanda Li, China

Chin-Teng Lin, Taiwan

Robert J. Marks II, USA
Erkki Oja, Finland
Nikhil R. Pal, India
Marios M. Polycarpou, USA
Leszek Rutkowski, Poland
Jennie Si, USA
Youxian Sun, China
Joos Vandewalle, Belgium
DeLiang Wang, USA
Fei-Yue Wang, USA
Donald C. Wunsch II, USA
Lei Xu, China
Xin Yao, UK
Gary G. Yen, USA
Bo Zhang, China
Nanning Zheng, China
Jacek M. Zurada, USA

Steering Committee Chairs

Jun Wang, Hong Kong
Derong Liu, China

Steering Committee Members

Jinde Cao, China
Shumin Fei, China
Chengan Guo, China
Min Han, China
Zeng-Guang Hou, China
Xiaofeng Liao, China
Bao-Liang Lu, China
Fuchun Sun, China
Zhang Yi, China
Fuliang Yin, China
Hujun Yin, UK
Huaguang Zhang, China
Jianwei Zhang, Germany

Organizing Committee Chairs

Hongwei Wang, China
Jianzhong Zhou, China
Yi Shen, China

Program Committee Chairs

Wen Yu, Mexico
Haibo He, USA
Nian Zhang, USA

Special Sessions Chairs

Sanqing Hu, USA
Youshen Xia, China
Yunong Zhang, China

Publications Chairs

Xiaolin Hu, China
Minghui Jiang, China
Qingshan Liu, China

Publicity Chairs

Tingwen Huang, Qatar
Paul S. Pang, New Zealand
Changyin Sun, China

Finance Chair

Xiaoping Wang, China

Registration Chairs

Charlie C. L. Wang, China
Zhenyuan Liu, China
Weifeng Zhu, China

Local Arrangements Chairs

Zhigang Zeng, China
Chao Qi, China
Liu Hong, China

Program Committee Members

José Alfredo, Brazil
Sabri Arik, Turkey
Xindi Cai, USA
Yu Cao, USA
Matthew Casey, UK
Emre Celebi, USA
Jonathan Chan, Thailand
Sheng Chen, UK
Yangquan Chen, USA
Ji-Xiang Du, China
Hai-Bin Duan, China
Andries Engelbrecht, South Africa
Péter érdi, USA
Jufeng Feng, China
Chaojin Fu, China
Wai Keung Fung, Canada
Erol Gelenbe, UK
Xinping Guan, China
Chengan Guo, China
Ping Guo, China
Qing-Long Han, Australia
Hanlin He, China
Daniel Ho, Hong Kong
Zhongsheng Hou, China
Huosheng Hu, UK
Jinglu Hu, Japan
Junhao Hu, China
Marc van Hulle, Belgium
Danchi Jiang, Australia
Haijun Jiang, China
Shunshoku Kanae, Japan
Rhee Man Kil, Republic of Korea
Sungshin Kim, Korea
Arto Klami, Finland
Rakhesh Singh Kshetrimayum, India
Hon Keung Kwan, Canada
Chuangdong Li, China
Kang Li, UK
Li Li, China
Michael Li, Australia
Ping Li, Hong Kong
Shutao Li, China
Xiaoli Li, UK
Xiaoou Li, Mexico
Yangmin Li, Macao
Hualou Liang, USA
Jinling Liang, China
Wudai Liao, China
Alan Liew, Australia
Ju Liu, China
Li Liu, USA
Meiqin Liu, China
Wenxin Liu, USA
Yan Liu, USA
Jianquan Lu, Hong Kong
Jinhu Lu, China
Wenlian Lu, China
Jinwen Ma, China
Ikuko Nishkawa, Japan
Seiichi Ozawa, Japan
Jaakko Peltonen, Finland
Juan Reyes, Mexico
Jose de Jesus Rubio, Mexico
Eng. Sattar B. Sadkhan, Iraq
Gerald Schaefer, UK
Michael Small, Hong Kong
Qiankun Song, China
Humberto Sossa, Mexico
Bingyu Sun, China
Norikazu Takahashi, Japan
Manchun Tan, China
Ying Tan, China
Christos Tjortjjs, UK
Michel Verleysen, Belgium
Bing Wang, UK
Dan Wang, China
Dianhui Wang, Australia
Meiqing Wang, China
Rubin Wang, China
Xin Wang, China
Zhongsheng Wang, China
Jinyu Wen, China
Wei Wu, China
Degui Xiao, China
Rui Xu, USA
Yingjie Yang, UK
Kun Yuan, China
Xiaoqin Zeng, China
Jie Zhang, UK
Liqing Zhang, China

Publications Committee Members

Guici Chen	Zhikun Wang
Huangqiong Chen	Shiping Wen
Shengle Fang	Ailong Wu
Lizhu Feng	Yongbo Xia
Junhao Hu	Li Xiao
Feng Jiang	Weina Yang
Bin Li	Zhanying Yang
Yanling Li	Tianfeng Ye
Mingzhao Li	Hongyan Yin
Lei Liu	Lingfa Zeng
Xiaoyang Liu	Yongchang Zhang
Cheng Wang	Yongqing Zhao
Xiaohong Wang	Song Zhu

Technical Committee Members

Helena Aidos	Shan Chen
Antti Ajanki,	Sheng Chen
Tholkappia AraSu	Siyue Chen
Hyeon Bae	TianYu Chen
Tao Ban	Wei Chen
Li Bin	Xi Chen
Binghuang Cai	Xiaochi Chen
Lingru Cai	Xiaofeng Chen
Xindi Cai	XinYu Chen
Qiao Cai	Xiong Chen
Chao Cao	Xuedong Chen
Hua Cao	Yongjie Chen
Jinde Cao	Zongzheng Chen
Kai Cao	Hao Cheng
Wenbiao Cao	Jian Cheng
Yuan Cao	Long Cheng
George Cavalcanti	Zunshui Cheng
Lei Chang	Rong Chu
Mingchun Chang	Bianca di Angeli C.S. Costa
Zhai Chao	Jose Alfredo Ferreira Costa
Cheng Chen	Dadian Dai
Gang Chen	Jianming Dai
Guici Chen	Jayanta Kumar Debnath
Ke Chen	Spiros Denaxas
Jiao Chen	Chengnuo Deng
Lei Chen	Gang Deng
Ming Chen	Jianfeng Deng
Rongzhang Chen	Kangfa Deng

Zhipo Deng	Li Hong
Xiaohua Ding	Liu Hong
Xiuzhen Ding	Ruibing Hou
Zhiqiang Dong	Cheng Hu
Jinran Du	Jin Hu
Hongwu Duan	Junhao Hu
Lijuan Duan	Hao Hu
Xiaopeng Duan	Hui Hu
Yasunori Endo	Ruibin Hu
Andries Engelbrecht	Sanqing Hu
Tolga Ensari	Xiaolin Hu
Zhengping Fan	Xiaoyan Hu
Fang Fang	Chi Huang
Haitao Fang	Darong Huang
Yuanda Fang	Diqiu Huang
June Feng	Dongliang Huang
Lizhu Feng	Gan Huang
Yunqing Feng	Huayong Huang
Avgoustinos Filippoupolitis	Jian Huang
Liang Fu	Li Huang
Ruhai Fu	Qifeng Huang
Fang Gao	Tingwen Huang
Lei Gao	Zhangcan Huang
Ruiling Gao	Zhenkun Huang
Daoyuan Gong	Zhilin Huang
Xiangguo Gong	Rey-Chue Hwang
Fanji Gu	Sae Hwang
Haibo Gu	Hui Ji
Xingsheng Gu	Tianyao Ji
Lihe Guan	Han Jia
Jun Guo	Danchi Jiang
Songtao Guo	Shaobo Jiang
Xu Guo	Wei Jiang
Fengqing Han	Wang Jiao
Pei Han	Xianfa Jiao
Qi Han	Yiannis Kanellopoulos
Weiwei Han	Wenjing Kang
Yishan Han	Anthony Karageorgos
Yunpeng Han	Masanori KaWakita
Hanlin He	Haibin Ke
Jinghui He	Seong-Joo Kim
Rui He	Peng Kong
Shan He	Zhanghui Kuang
Tonejun He	Lingcong Le
Tongjun He	Jong Min Lee
Wangli He	Liu Lei
Huosheng Hu	Siyu Leng

Bing Li
Changping Li
Chuandong Li
Hui Li
Jian Li
Jianmin Li
Jianxiang Li
Kelin Li
Kezan Li
Lei Li
Li Li
Liping Li
Lulu Li
Ming Li
Na Li
Ping Li
Qi Li
Song Li
Wei qun Li
Wenlong Li
Wentian Li
Shaokang Li
Shiying Li
Tian Li
Wei Li
Wu Li
Xiang Li
Xiaoli Li
Xiaoou Li
Xin Li
Xinghai Li
Xiumin Li
Yanlin Li
Yanling Li
Yong Li
Yongfei Li
Yongmin Li
Yuechao Li
Zhan Li
Zhe Li
Jinling Liang
Wudai Liao
Wei Lin
Zhihao Lin
Yunqing Ling
Alex Liu
Bo Liu
Da Liu
Dehua Li
Dayuan Liu
Dongbing Liu
Desheng Liu
F. C. Liu
Huaping Liu
Jia Liu
Kangqi Liu
Li Liu
Ming Liu
Qian Liu
Qingshan Liu
Shangjin Liu
Shenquan Liu
Shi Liu
Weiqi Liu
Xiaoyang Liu
Xiuquan Liu
Xiwei Liu
XinRong Liu
Yan Liu
Yang Liu
Yawei Liu
Yingju Liu
Yuxi Liu
Zhenyuan Liu
Zijian Liu
Yimin Long
Georgios Loukas
Jinhu Lu
Jianquan Lu
Wen Lu
Wenlian Lu
Wenqian Lu
Tongting Lu
Qiuming Luo
Xucheng Luo
Chaohua Ma
Jie Ma
Liefeng Ma
Long Ma
Yang Ma
Zhiwei Ma
Xiaoou Mao
Xuehui Mei
Xiangpei Meng

Xiangyu Meng
Zhaohui Meng
Guo Min
Rui Min
Yuanneng Mou
Junichi Murata
Puyan Nie
Xiushan Nie
Gulay Oke
Ming Ouyang
Yao Ouyang
Seiichi Ozawa
Neyir Ozcan
Joni Pajarinen
Hongwei Pan
Linqiang Pan
Yunpeng Pan
Tianqi Pang
Kyungseo Park
Xiaohan Peng
Zaiyun Peng
Gao Pingan
Liquan Qiu
Jianlong Qiu
Tapani Raiko
Congjun Rao
Fengli Ren
Jose L. Rosseilo
Gongqin Ruan
Quan Rui
Sattar B. Sadkhan
Renato Jose Sassi Sassi
Sibel Senan
Sijia Shao
Bo Shen
Enhua Shen
Huayu Shen
Meili Shen
Zifei Shen
Dianyang Shi
Jinrui Shi
Lisha Shi
Noritaka Shigei
Atsushi Shimada
Jiaqi Song
Wen Song
Yexin Song
Zhen Song
Zhu Song
Gustavo Fontoura de Souza
Kuo-Ho Su
Ruiqi Su
Cheng Sun
Dian Sun
Junfeng Sun
Lisha Sun
Weipeng Sun
Yonghui Sun
Zhaowan Sun
Zhendong Sun
Manchun Tan
Xuehong Tan
Yanxing Tan
Zhiguo Tan
Bing Tang
Hao Tang
Yili Tang
Gang Tian
Jing Tian
Yuguang Tian
Stelios Timotheou
Shozo Tokinaga
Jun Tong
Joaquin Torres Sospedra
Hiroshi Wakuya
Jin Wan
B.H. Wang
Cheng Wang
Fan Wang
Fen Wang
Gang Wang
Gaoxia Wang
Guanjun Wang
Han Wang
Heding Wang
Hongcui Wang
Huayong Wang
Hui Wang
Huiwei Wang
Jiahai Wang
Jian Wang
Jin Wang
Juzhi Wang
Kai Wang

Lan Wang	Zhiguo Xia
Lili Wang	Xun Xiang
Lu Wang	Chengcheng Xiao
Qilin Wang	Donghua Xiao
Qingyun Wang	Jiangwen Xiao
Suqin Wang	Yongkang Xiao
Tian Wang	Yonkang Xiao
Tianxiong Wang	Yong Xie
Tonghua Wang	Xiaofei Xie
Wei Wang	Peng Xin
Wenjie Wang	Chen Xiong
Xiao Wang	Jinghui Xiong
Xiaoping Wang	Wenjun Xiong
Xiong Wang	Anbang Xu
Xudong Wang	Chen Xu
Yang Wang	Hesong Xu
Yanwei Wang	Jianbing Xu
Yao Wang	Jin Xu
Yiping Wang	Lou Xu
Yiyu Wang	Man Xu
Yue Wang	Xiufen Yu
Zhanshan Wang	Yan Xu
Zhengxia Wang	Yang Xu
Zhibo Wang	Yuanlan Xu
Zhongsheng Wang	Zhaodong Xu
Zhihui Wang	Shujing Yan
Zidong Wang	Dong Yang
Zhuo Wang	Fan Yang
Guoliang Wei	Gaobo Yang
Li Wei	Lei Yang
Na Wei	Sihai Yang
Shuang Wei	Tianqi Yang
Wenbiao Wei	Xiaolin Yang
Yongchang Wei	Xing Yang
Xiaohua Wen	Xue Yang
Xuexin Wen	Yang Yang
Junmei Weng	Yongqing Yang
Yixiang Wu	Yiwen Yang
You Wu	Hongshan Yao
Huaiqin Wu	John Yao
Zhihai Wu	Xianfeng Ye
Bin Xia	Chenfu Yi
Weiguo Xia	Aihua Yin
Yonghui Xia	Lewen Yin
Youshen Xia	Qian Yin
Zhigu Xia	Yu Ying

Xu Yong
 Yuan You
 Shuai You
 Chenglong Yu
 Liang Yu
 Lin Yu
 Liqiang Yu
 Qing Yu
 Yingzhong Yu
 Zheyi Yu
 Jinhui Yuan
 Peijiang Yuan
 Eylem Yucel
 Si Yue
 Jianfang Zeng
 Lingjun Zeng
 Ming Zeng
 Yi Zeng
 Zeyu Zhang
 Zhigang Zeng
 Cheng Zhang
 Da Zhang
 Hanling Zhang
 Haopeng Zhang
 Kaifeng Zhang
 Jiakai Zhang
 Jiajia Zhang
 Jiangjun Zhang
 Jifan Zhang
 Jinjian Zhang
 Liming Zhang
 Long Zhang
 Qi Zhang
 Rui Zhang
 Wei Zhang
 Xiaochun Zhang
 Xiong Zhang
 Xudong Zhang
 Xuguang Zhang
 Yang Zhang
 Yangzhou Zhang
 Yinxue Zhang
 Yunong Zhang
 Zhaoxiong Zhang

YuanYuan
 Bin Zhao
 Jin Zhao
 Le Zhao
 Leina Zhao
 Qibin Zhao
 Xiaquan Zhao
 Zhenjiang Zhao
 Yue Zhen
 Changwei Zheng
 Huan Zheng
 Lina Zheng
 Meijun Zheng
 Quanchao Zheng
 Shitao Zheng
 Ying Zheng
 Xun Zheng
 Lingfei Zhi
 Ming Zhong
 Benhai Zhou
 Jianxiang Zhou
 Jiao Zhou
 Jin Zhou
 Jinnong Zhou
 Junming Zhou
 Lin Zhou
 Rong Zhou
 Song Zhou
 Xiang Zhou
 Xiuling Zhou
 Yiduo Zhou
 Yinlei Zhou
 Yuan Zhou
 Zhenqiao Zhou
 Ze Zhou
 Zhouliu Zhou
 Haibo Zhu
 Ji Zhu
 Jiajun Zhu
 Tanyuan Zhu
 Zhenqian Zhu
 Song Zhu
 Xunlin Zhu
 Zhiqiang Zuo

Table of Contents – Part III

Optimization

A Modified Projection Neural Network for Linear Variational Inequalities and Quadratic Optimization Problems	1
<i>Minghui Jiang, Yongqing Zhao, and Yi Shen</i>	
Diversity Maintenance Strategy Based on Global Crowding	10
<i>Qiong Chen, Shengwu Xiong, and Hongbing Liu</i>	
Hybrid Learning Enhancement of RBF Network Based on Particle Swarm Optimization	19
<i>Sultan Noman Qasem and Siti Mariyam Shamsuddin</i>	
Chaos Cultural Particle Swarm Optimization and Its Application	30
<i>Ying Wang, Jianzhong Zhou, Youlin Lu, Hui Qin, and Yongchuan Zhang</i>	
Application of Visualization Method to Concrete Mix Optimization	41
<i>Bin Shi, Liexiang Yan, and Quan Guo</i>	
A Novel Nonparametric Regression Ensemble for Rainfall Forecasting Using Particle Swarm Optimization Technique Coupled with Artificial Neural Network	49
<i>Jiansheng Wu and Enhong Chen</i>	
A Revised Neural Network for Solving Quadratic Programming Problems	59
<i>Yinjie Sun</i>	
The Separation Property Enhancement of Liquid State Machine by Particle Swarm Optimization	67
<i>Jiangshuai Huang, Yongji Wang, and Jian Huang</i>	
A Class of New Large-Update Primal-Dual Interior-Point Algorithms for $P_*(\kappa)$ Linear Complementarity Problems	77
<i>Huaping Chen, Mingwang Zhang, and Yuqin Zhao</i>	
A Novel Artificial Immune System for Multiobjective Optimization Problems	88
<i>Jiaquan Gao and Lei Fang</i>	
A Neural Network Model for Solving Nonlinear Optimization Problems with Real-Time Applications	98
<i>Alaeddin Malek and Maryam Yashtini</i>	

Evolutionary Markov Games Based on Neural Network	109
<i>Liu Weibing, Wang Xianjia, and Huang Binbin</i>	
Another Simple Recurrent Neural Network for Quadratic and Linear Programming	116
<i>Xiaolin Hu and Bo Zhang</i>	
A Particle Swarm Optimization Algorithm Based on Genetic Selection Strategy	126
<i>Qin Tang, Jianyou Zeng, Hui Li, Changhe Li, and Yong Liu</i>	
Structure Optimization Algorithm for Radial Basis Probabilistic Neural Networks Based on the Moving Median Center Hyperspheres Algorithm	136
<i>Ji-Xiang Du and Chuan-Min Zhai</i>	
Nonlinear Component Analysis for Large-Scale Data Set Using Fixed-Point Algorithm	144
<i>Weiya Shi and Yue-Fei Guo</i>	
Optimal Reactive Power Dispatch Using Particle Swarms Optimization Algorithm Based Pareto Optimal Set	152
<i>Yan Li, Pan-pan Jing, De-feng Hu, Bu-han Zhang, Cheng-xiong Mao, Xin-bo Ruan, Xiao-yang Miao, and De-feng Chang</i>	
Robotics	
A Robust Non-Line-Of-Sight Error Mitigation Method in Mobile Position Location	162
<i>Sumei Chen, Ju Liu, and Lin Xue</i>	
Research on SSVEP-Based Controlling System of Multi-DoF Manipulator	171
<i>Hui Shen, Li Zhao, Yan Bian, and Longteng Xiao</i>	
Tracking Control of Robot Manipulators via Orthogonal Polynomials Neural Network	178
<i>Hongwei Wang and Shuanghe Yu</i>	
Q-Learning Based on Dynamical Structure Neural Network for Robot Navigation in Unknown Environment	188
<i>Junfei Qiao, Ruiyuan Fan, Honggui Han, and Xiaogang Ruan</i>	
Research on Mobile Robot’s Motion Control and Path Planning	197
<i>Shigang Cui, Xuelian Xu, Li Zhao, Liguo Tian, and Genghuang Yang</i>	

A New Cerebellar Model Articulation Controller for Rehabilitation Robots	207
<i>Shan Liu, Yongji Wang, Yongle Xie, Shuyan Jiang, and Jinsong Meng</i>	
Layer-TERRAIN: An Improved Algorithm of TERRAIN Based on Sequencing the Reference Nodes in UWSNs	217
<i>Yue Liang and Zhong Liu</i>	
A Hybrid Neural Network Method for UAV Attack Route Integrated Planning	226
<i>Nan Wang, Xueqiang Gu, Jing Chen, Lincheng Shen, and Min Ren</i>	
Hybrid Game Theory and D-S Evidence Approach to Multiple UCAVs Cooperative Air Combat Decision	236
<i>Xingxing Wei, Haibin Duan, and Yanran Wang</i>	
FCMAC Based Guidance Law for Lifting Reentry Vehicles	247
<i>Hao Wu, Chuanfeng Li, and Yongji Wang</i>	
Hybrid Filter Based Simultaneous Localization and Mapping for a Mobile Robot	257
<i>Kyung-Sik Choi, Bong-Keun Song, and Suk-Gyu Lee</i>	
Using Toe-off Impulse to Control Chaos in the Simplest Walking Model via Artificial Neural Network	267
<i>Saeed Jamali, Karim Faez, Sajjad Taghvaei, and Mostafa Ozlati Moghadam</i>	
Reinforcement Learning Control of a Real Mobile Robot Using Approximate Policy Iteration	278
<i>Pengcheng Zhang, Xin Xu, Chunming Liu, and Qiping Yuan</i>	

Image Processing

A Simple Neural Network for Enhancement of Image Acuity by Fixational Instability	289
<i>Daqing Yi, Ping Jiang, and Jin Zhu</i>	
A General-Purpose FPGA-Based Reconfigurable Platform for Video and Image Processing	299
<i>Jie Li, Haibo He, Hong Man, and Sachi Desai</i>	
Image Analysis by Modified Krawtchouk Moments	310
<i>Luo Zhu, Jiaping Liao, Xiaoqin Tong, Li Luo, Bo Fu, and Guojun Zhang</i>	

Efficient Provable Secure ID-Based Directed Signature Scheme without Random Oracle	318
<i>Jianhong Zhang, Yixian Yang, and Xinxin Niu</i>	
Mask Particle Filter for Similar Objects Tracking	328
<i>Huaping Liu, Fuchun Sun, and Meng Gao</i>	
An Efficient Wavelet Based Feature Extraction Method for Face Recognition	337
<i>Iman Makaremi and Majid Ahmadi</i>	
Face Recognition Based on Histogram of Modular Gabor Feature and Support Vector Machines	346
<i>Xiaodong Li, Shumin Fei, and Tao Zhang</i>	
Feature-Level Fusion of Iris and Face for Personal Identification	356
<i>Zhifang Wang, Qi Han, Xiamu Niu, and Christoph Busch</i>	
Watermark Image Restoration Method Based on Block Hopfield Network	365
<i>Xiaohong Ma, Xin Li, and Hualou Liang</i>	
An English Letter Recognition Algorithm Based Artificial Immune	371
<i>Chunlin Liang, Lingxi Peng, Yindie Hong, and Jing Wang</i>	
Interpretation of Ambiguous Zone in Handwritten Chinese Character Images Using Bayesian Network	380
<i>Zhongsheng Cao, Zhewen Su, and Yuanzhen Wang</i>	
Weather Recognition Based on Images Captured by Vision System in Vehicle	390
<i>Xunshi Yan, Yupin Luo, and Xiaoming Zheng</i>	
Selecting Regions of Interest for the Diagnosis of Alzheimer Using Brain SPECT Images	399
<i>Diego Salas-Gonzalez, Juan M. Górriz, Javier Ramírez, Ignacio Álvarez, Míriam López, Fermín Segovia, and Carlos G. Puntonet</i>	
Face Image Recognition Combining Holistic and Local Features	407
<i>Chen Pan and Feilong Cao</i>	
3D Representative Face and Clustering Based Illumination Estimation for Face Recognition and Expression Recognition	416
<i>Zheng Zhang, Zheng Zhao, and Gang Bai</i>	
Bilateral Two-Dimensional Locality Preserving Projections with Its Application to Face Recognition	423
<i>Xiao-Guo Wang</i>	

DT-CWT Feature Structure Representation for Face Recognition under Varying Illumination Using EMD <i>Yuehui Sun and Di Zhang</i>	429
Spatially Smooth Subspace Face Recognition Using LOG and DOG Penalties <i>Wangmeng Zuo, Lei Liu, Kuanquan Wang, and David Zhang</i>	439
Nonnegative-Least-Square Classifier for Face Recognition <i>Nhat Vo, Bill Moran, and Subhash Challa</i>	449
A Novel Model for Recognition of Compounding Nouns in English and Chinese <i>Lishu Li, Jiawei Chen, Qinghua Chen, and Fukang Fang</i>	457
Orthogonal Quadratic Discriminant Functions for Face Recognition <i>Suicheng Gu, Ying Tan, and Xingui He</i>	466
LISA: Image Compression Scheme Based on an Asymmetric Hierarchical Self-Organizing Map <i>Cheng-Fa Tsai and Yu-Jiun Lin</i>	476
A Method of Human Skin Region Detection Based on PCNN <i>Lijuan Duan, Zhiqiang Lin, Jun Miao, and Yuanhua Qiao</i>	486
An Adaptive Hybrid Filtering for Removing Impulse Noise in Color Images <i>Xuan Guo, Baoping Guo, Tao Hu, and Ou Yang</i>	494
A Multi-Stage Neural Network Model for Human Color Vision <i>Charles Q. Wu</i>	502
Lead Field Space Projection for Spatiotemporal Imaging of Independent Brain Activities <i>Huilin Chan, Yong-Sheng Chen, Li-Fen Chen, Tzu-Hua Chen, and I-Tzu Chen</i>	512
Morphological Hetero-Associative Memories Applied to Restore True-Color Patterns <i>Roberto A. Vázquez and Humberto Sossa</i>	520
Signal Processing	
A Novel Method for Analyzing Dynamic Complexity of EEG Signals Using Symbolic Entropy Measurement <i>Lisha Sun, Jun Yu, and Patch J. Beadle</i>	530

Phase Self-amending Blind Equalization Algorithm Using Feedforward Neural Network for High-Order QAM Signals in Underwater Acoustic Channels	538
<i>Yasong Luo, Zhong Liu, Pengfei Peng, and Xuezhi Fu</i>	
An Adaptive Channel Handoff Strategy for Opportunistic Spectrum Sharing in Cognitive Global Control Plane Architecture	546
<i>Zhiming Xu, Yu Wang, Jingguo Zhu, and Jian Tang</i>	
A Generalization of the Bent-Function Sequence Construction	557
<i>Yongbo Xia, Yan Sui, and Junhao Hu</i>	
An Efficient Large-Scale Volume Data Compression Algorithm	567
<i>Degui Xiao, Liping Zhao, Lei Yang, Zhiyong Li, and Kenli Li</i>	
Simultaneous Synchronization of Text and Speech for Broadcast News Subtitling	576
<i>Jie Gao, Qingwei Zhao, Ta Li, and Yonghong Yan</i>	
A Perceptual Weighting Filter Based on ISP Pseudo-cepstrum and Its Application in AMR-WB	586
<i>Fenglian Li and Xueying Zhang</i>	
Video Fingerprinting by Using Boosted Features	596
<i>Huicheng Lian and Jing Xu</i>	
Reference Signal Impact on EEG Energy	605
<i>Sanqing Hu, Matt Stead, Hualou Liang, and Gregory A. Worrell</i>	
Multichannel Blind Deconvolution Using the Conjugate Gradient	612
<i>Bin Xia</i>	
An Improvement of HSMM-Based Speech Synthesis by Duration-Dependent State Transition Probabilities	621
<i>Jing Tao and Wenju Liu</i>	
Biomedical Applications	
Handprint Recognition: A Novel Biometric Technology	630
<i>Guiyu Feng, Qi Zhao, Miyi Duan, Dewen Hu, and Yabin Hu</i>	
Single Trial Evoked Potentials Estimation by Using Wavelet Enhanced Principal Component Analysis Method	638
<i>Ling Zou, Zhenghua Ma, Shuyue Chen, Suolan Liu, and Renlai Zhou</i>	
Fourier Volume Rendering on GPGPU	648
<i>Degui Xiao, Yi Liu, Lei Yang, Zhiyong Li, and Kenli Li</i>	

An Improved Population Migration Algorithm for the Prediction of Protein Folding	657
<i>Huafeng Chen and Jianyong Wang</i>	
Gene Sorting in Differential Evolution	663
<i>Remi Tassing, Desheng Wang, Yongli Yang, and Guangxi Zhu</i>	
Enhancement of Chest Radiograph Based on Wavelet Transform	675
<i>Zhenghao Shi, Lifeng He, Tsuyoshi Nakamura, and Hidenori Itoh</i>	
Application of DNA Computing by Self-assembly on 0-1 Knapsack Problem	684
<i>Guangzhao Cui, Cuijing Li, Xuncai Zhang, Yanfeng Wang, Xinbo Qi, Xiaoguang Li, and Haobin Li</i>	
Learning Kernel Matrix from Gene Ontology and Annotation Data for Protein Function Prediction	694
<i>Yiming Chen, Zhoujun Li, and Junwan Liu</i>	
Improved Quantum Evolutionary Algorithm Combined with Chaos and Its Application	704
<i>Jianhua Xiao</i>	

Fault Diagnosis

Fault Diagnosis of Nonlinear Analog Circuits Using Neural Networks and Multi-Space Transformations	714
<i>Yigang He and Wenji Zhu</i>	
An Intelligent Fault Diagnosis Method Based on Multiscale Entropy and SVMs	724
<i>Long Zhang, Guoliang Xiong, Hesheng Liu, Huijun Zou, and Weizhong Guo</i>	
Multi-objective Robust Fault Detection Filter Design in a Finite Frequency Range	733
<i>Yu Cui, Xin-han Huang, and Min Wang</i>	
Intelligent Technique and Its Application in Fault Diagnosis of Locomotive Bearing Based on Granular Computing	744
<i>Zhang Zhousuo, Yan Xiaoxu, and Cheng Wei</i>	
Analysis of Two Neural Networks in the Intelligent Faults Diagnosis of Metallurgic Fan Machinery	755
<i>Jiangang Yi and Peng Zeng</i>	
Research on the Diagnosis of Insulator Operating State Based on Improved ANFIS Networks	762
<i>Zipeng Zhang, Shuqing Wang, Liqin Xue, and Xiaohui Yuan</i>	

Fault Diagnosis of Analog IC Based on Wavelet Neural Network Ensemble 772
Lei Zuo, Ligang Hou, Wuchen Wu, Jinhui Wang, and Shuqin Geng

Dynamic Neural Network-Based Fault Detection and Isolation for Thrusters in Formation Flying of Satellites 780
Arturo Valdes, K. Khorasani, and Liying Ma

Passivity Analysis of a General Form of Recurrent Neural Network with Multiple Delays 794
Jinhua Huang and Jiqing Liu

Comparative Analysis of Corporate Failure Prediction Methods: Evidence from Chinese Firms 801
Haicong Yang

Telecommunication, Sensor Network and Transportation Systems

An Adaline-Based Location Algorithm for Wireless Sensor Network 809
Fengjun Shang

Remote Estimation with Sensor Scheduling 819
Li Xiao, Zigang Sun, Desen Zhu, and Mianyun Chen

An Improved Margin Adaptive Subcarrier Allocation with Fairness for Multiuser OFDMA System 829
Tan Li, Gang Su, Guangxi Zhu, Jun Jiang, and Hui Zhang

Detecting Community Structure in Networks by Propagating Labels of Nodes 839
Chuanjun Pang, Fengjing Shao, Rencheng Sun, and Shujing Li

Algorithm for Multi-sensor Asynchronous Track-to-Track Fusion 847
Cheng Cheng and Jinfeng Wang

Remote Sensing Based on Neural Networks Model for Hydrocarbon Potentials Evaluation in Northeast China 855
Shengbo Chen

A Multiple Weighting Matrices Selection Scheme Based on Orthogonal Random Beamforming for MIMO Downlink System 864
Li Tan, Gang Su, Guangxi Zhu, and Peng Shang

A Novel Adaptive Reclosure Criterion for HV Transmission Lines Based on Wavelet Packet Energy Entropy 874
Yuanyuan Zhang, Qingwu Gong, and Xi Shi

Pre-estimate on Transport Volume of Container in Xiangjiang Catchment	882
<i>Jian-Lan Zhou</i>	
RTKPS: A Key Pre-distribution Scheme Based on Rooted-Tree in Wireless Sensor and Actor Network	890
<i>Zhicheng Dai, Zhi Li, Bingwen Wang, and Qiang Tang</i>	
Urban Road Network Modeling and Real-Time Prediction Based on Householder Transformation and Adjacent Vector	899
<i>Shuo Deng, Jianming Hu, Yin Wang, and Yi Zhang</i>	
Research on Method of Double-Layers BP Neural Network in Prediction of Crossroads' Traffic Volume	909
<i>Yuming Mao, Shiyiing Shi, Hai Yang, and Yuanyuan Zhang</i>	
Design and Implementation of the Structure Health Monitoring System for Bridge Based on Wireless Sensor Network	915
<i>An Yin, Bingwen Wang, Zhuo Liu, and Xiaoya Hu</i>	
Saving Energy in Wireless Sensor Networks Based on Echo State Networks	923
<i>Ling Qin, Rongqiang Hu, and Qi Zhang</i>	
Enlargement of Measurement Range in a Fiber-Optic Ice Sensor by Artificial Neural Network	929
<i>Wei Li, Jie Zhang, Ying Zheng, and Lin Ye</i>	
Epidemic Spreading with Variant Infection Rates on Scale-Free Network	937
<i>Liu Hong, Min Ouyang, Zijun Mao, and Xueguang Chen</i>	
Interdependency Analysis of Infrastructures	948
<i>Zijun Mao, Liu Hong, Qi Fei, and Ming OuYang</i>	
Back Propagation Neural Network Based Lifetime Analysis of Wireless Sensor Network	956
<i>Wenjun Yang, Bingwen Wang, Zhuo Liu, and Xiaoya Hu</i>	
Applications I	
Estimation of Rock Mass Rating System with an Artificial Neural Network	963
<i>Zhi Qiang Zhang, Qing Ming Wu, Qiang Zhang, and Zhi Chao Gong</i>	
Comparative Study on Three Voidage Measurement Methods for Two-Phase Flow	973
<i>Youmin Guo and Zhenrui Peng</i>	

A New Approach to Improving ICA-Based Models for the Classification of Microarray Data	983
<i>Kun-Hong Liu, Bo Li, Jun Zhang, and Ji-Xiang Du</i>	
Multiple Trend Breaks and Unit Root Hypothesis: Empirical Evidence from China's GDP(1952-2006)	993
<i>Shusheng Li and Zhao-hui Liang</i>	
An Adaptive Wavelet Networks Algorithm for Prediction of Gas Delay Outburst	1000
<i>Xinyu Li</i>	
Traffic Condition Recognition of Probability Neural Network Based on Floating Car Data	1007
<i>Gengqi Guo, Chengtao Cao, Jiuzhong Li, and Shuo Shi</i>	
Combined Neural Network Approach for Short-Term Urban Freeway Traffic Flow Prediction	1017
<i>Ruimin Li and Huapu Lu</i>	
Facial Expression Recognition in Video Sequences	1026
<i>Shenchuan Tai and Hungfu Huang</i>	
An AFSA-TSGM Based Wavelet Neural Network for Power Load Forecasting	1034
<i>Dongxiao Niu, Zhihong Gu, and Yunyun Zhang</i>	
Comparative Analyses of Computational Intelligence Models for Load Forecasting: A Case Study in the Brazilian Amazon Power Suppliers . . .	1044
<i>Liviane P. Rego, Ádamo L. de Santana, Guilherme Conde, Marcelino S. da Silva, Carlos R.L. Francês, and Cláudio A. Rocha</i>	
An Efficient and Robust Algorithm for Improving the Resolution of Video Sequences	1054
<i>Yubing Han, Rushan Chen, and Feng Shu</i>	
Research on Variable Step-Size Blind Equalization Algorithm Based on Normalized RBF Neural Network in Underwater Acoustic Communication	1063
<i>Xiaoling Ning, Zhong Liu, and Yasong Luo</i>	
The Analysis of Aircraft Maneuver Efficiency within Extend Flight Envelop	1071
<i>Hao Long and Shujie Song</i>	
Application of BP Neural Network in Stock Market Prediction	1082
<i>Bin Fang and Shoufeng Ma</i>	

A Research of Physical Activity's Influence on Heart Rate Using Feedforward Neural Network	1089
<i>Feng Xiao, Ming Yuchi, Jun Jo, Ming-yue Ding, and Wen-guang Hou</i>	
Bi-directional Prediction between Weld Penetration and Processing Parameters in Electron Beam Welding Using Artificial Neural Networks	1097
<i>Xianfeng Shen, Wenrong Huang, Chao Xu, and Xingjun Wang</i>	
Analysis of Nonlinear Dynamic Structure for the Shanghai Stock Exchange Index	1106
<i>Yu Dong and Hu Song</i>	
A Direct Approach to Achieving Maximum Power Conversion in Wind Power Generation Systems	1112
<i>Y.D. Song, X.H. Yin, Gary Lebby, and Liguao Weng</i>	
Applications II	
Synthetic Modeling and Policy Simulation of Regional Economic System: A Case Study	1122
<i>Zhi Yang, Wei Zeng, Hongtao Zhou, Lingru Cai, Guangyong Liu, and Qi Fei</i>	
Industrial Connection Analysis and Case Study Based on Theory of Industrial Gradient	1130
<i>Zhi Yang, Wei Zeng, Hongtao Zhou, Ying Li, and Qi Fei</i>	
Extracting Schema from Semistructured Data with Weight Tag	1137
<i>Jiuzhong Li and Shuo Shi</i>	
Designing Domain Work Breakdown Structure (DWBS) Using Neural Networks	1146
<i>Yongjun Bai, Yong Zhao, Yang Chen, and Lu Chen</i>	
Practical Hardware Implementation of Self-configuring Neural Networks	1154
<i>Josep L. Rosselló, Vincent Canals, Antoni Morro, and Ivan de Paül</i>	
Research on Multi-Agent Parallel Computing Model of Hydrothermal Economic Dispatch in Power System	1160
<i>Bu-han Zhang, Junfang Li, Yan Li, Chengxiong Mao, Xin-bo Ruan, and Jianhua Yang</i>	
Fast Decoupled Power Flow Using Interval Arithmetic Considering Uncertainty in Power Systems	1171
<i>Shouxiang Wang, Chengshan Wang, Gaolei Zhang, and Ge Zhao</i>	

Power System Aggregate Load Area Dynamic Modeling by Learning Based on WAMS	1179
<i>Huimin Yang and Jinyu Wen</i>	
Optimal Preventive Maintenance Inspection Period on Reliability Improvement with Bayesian Network and Hazard Function in Gantry Crane	1189
<i>Gyeondong Baek, Kangkil Kim, and Sungshin Kim</i>	
Application of RBF Network Based on Immune Algorithm to Predicting of Wastewater Treatment	1197
<i>Hongtao Ye, Fei Luo, and Yuge Xu</i>	
HLA-Based Emergency Response Plan Simulation and Practice over Internet	1203
<i>Wan Hu, Hong Liu, and Qing Yang</i>	
Dynamic Cooperation Mechanism in Supply Chain for Perishable Agricultural Products under One-to-Multi	1212
<i>Lijuan Wang, Xichao Sun, and Feng Dang</i>	
Primary Research on Urban Mass Panic Based on Computational Methods for Experiments	1222
<i>Xi Chen, Qi Fei, and Wei Li</i>	
Virtual Reality Based Nuclear Steam Generator Ageing and Life Management Systems	1230
<i>Yajin Liu, Jiang Guo, Peng Liu, Lin Zhou, and Jin Jiang</i>	
Author Index	1241

A Modified Projection Neural Network for Linear Variational Inequalities and Quadratic Optimization Problems

Minghui Jiang¹, Yongqing Zhao¹, and Yi Shen²

¹ Institute of Nonlinear and Complex System, China Three Gorges University, YiChang, Hubei 443002, China

² Department of Control Science and Engineering, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China

Abstract. Variational inequalities provide us with a tool to study a wide class of optimization arising in pure and applied sciences. In the paper, we present a neural network for solving linear variational inequalities and quadratic optimization by using a projection techniques. We also consider the global uniqueness of the solution of the neural network as well as the convergence of the modified projection neural network. Our results present a significant improvement of previously known projection methods for solving variational inequalities and related optimization problems. Two simulation examples are provided to show the effectiveness of the approach and applicability of the proposed criteria.

Keywords: Neural network, Stability, Optimization.

1 Introduction

Variational inequalities, which was studied in the sixties, has emerged as an interesting branch of applicable mathematics with a wide range of applications in industry, finance, and applied sciences. While Optimization problems, such as quadratic programming problems, are special cases of variational inequalities. Meanwhile Optimization problems arise in a wide variety of scientific and engineering applications including signal processing, function approximation, regression analysis, and so on. In many engineering and scientific applications, the real-time solution of optimization problems and variational inequality are widely required. However, traditional algorithms for digital computers may not be efficient since the computing time required for a solution is greatly dependent on the dimension and structure of the problems. One possible and very promising approach to real-time optimization and variational inequality is to apply artificial neural networks. Because of the inherent massive parallelism, the neural network approach can solve optimization problems and variational inequality in running time at the orders of magnitude much faster than those of the most popular algorithms executed on general-purpose digital computers. The introduction of artificial neural networks in optimization and variational inequality

was stated in 1980s . Since then, significant research results have been achieved for various optimization and variational inequality. The neural network for solving programming problems was first proposed by Tank and Hopfield [1] in 1986. In 1987, Kennedy and Chua [2] proposed an improved model that always guaranteed convergence. However, their new model converges to only an approximation of the optimal solution. In 1990, Rodriguez-Vazquez et al. [3] presented a class of neural networks for solving optimization problems. Based on dual and projection methods [4] Xia et al. [5-9] presented several neural network for solving variational inequality and quadratic programming problems. Zhang [10] proposed the Lagrangian network and Wang et al. Chen and Fang [11] proposed a delayed neural network for solving convex quadratic programming problems by using the penalty function approach. However, this network can not converge an exact optimal solution and has an implementation problem when the penalty parameter is very large. In [12], the linear optimization neural network for associative memory was proposed by Tao, Liu and Cui. Effati and Nazemi [13] proposed a class of neural networks for solving optimization problems, but the conditions in the paper are hard to verify. The projection neural network for solving monotone linear variational inequalities and linear and quadratic optimization was given by Hu and Wang in [14], and the some interesting results are obtained. To speed the convergence of the projection neural network for solving the linear variational inequalities and quadratic optimization, we propose the modified projection neural network. Compared with the simulation results of the neural networks in other references, the convergence rate in this paper is more than one in [15].

This paper consists of the following sections. Section 2 describes some preliminaries. Existence, uniqueness and exponential stability of the modified projected neural network are discussed in the Section 3. In Section 4, two illustrative examples are given to verify the effectiveness of the results in this paper. Finally, concluding remarks are made in Section 5.

2 Preliminaries

In this paper, we are concerned with the following linear variational inequality(LVI): find $x^* \in \Omega$ such that

$$(Qx^* + c)^T(x - x^*) \geq 0 \quad \forall x \in \Omega \quad (1)$$

where $Q \in R^{n \times n}$, and $c \in R^n$, and

$$\Omega = \{x \in R^n \mid d \leq x \leq h\} \quad (2)$$

with $d, h \in R^n$ are bounded.

The above LVI is equivalent to the following quadratic programming problem[15]

$$\begin{aligned} & \text{minimize} && f(x) = \frac{1}{2}x^T Qx + c^T x \\ & \text{subject to} && x \in \Omega \end{aligned} \quad (3)$$

where the parameters are the same as in (II) and (I8). Recently, Xia and Wang [5,8], Hu and Wang [14] designed the projection neural networks for solving the above variational inequality and quadratic optimization problem.

To solve (II) and (I7), the following modified projection neural network model was considered in [15].

$$\frac{dx}{dt} = -x + P_{\Omega}(x - \alpha \nabla f(x)), \quad \forall x \in R^n \quad (4)$$

where $P_{\Omega}(x) = \{P_{\Omega}(x_1), P_{\Omega}(x_2), \dots, P_{\Omega}(x_n)\}^T$, $\nabla f(x) = Qx + c$, $\alpha > 0$, $\nabla^2 f(x) = Q$ and

$$P_{\Omega}(x_i) = \begin{cases} d_i & x_i < d_i \\ x_i & d_i \leq x_i \leq h_i \\ h_i & x_i > h_i \end{cases} \quad (5)$$

If x is an equilibrium point of the neural network (4), then

$$x = P_{\Omega}(x - \alpha \nabla f(x)).$$

Thus,

$$x = P_{\Omega}(-\alpha \nabla f(x) + P_{\Omega}(x - \alpha \nabla f(x))).$$

Therefore, we can consider the following modified projection neural network

$$\frac{dx}{dt} = -x + P_{\Omega}(-\alpha \nabla f(P_{\Omega}(x - \alpha \nabla f(x))) + P_{\Omega}(x - \alpha \nabla f(x))), \quad \forall x \in R^n \quad (6)$$

Remark 1. The term x in $P_{\Omega}(x - \alpha \nabla f(x))$ in the projection neural network (4) is replaced with $P_{\Omega}(x - \alpha \nabla f(x))$ as "predictor step" in the modified projection neural network (6).

For the convenience of discussion, some notations and definitions are introduced. Square matrix $Q > 0$ denotes by positive definite, and $\|x\|$ means the l_2 norm of a vector x ; $\lambda_{min}(Q)$, $\lambda_{max}(Q)$ stands for the minimum and maximum eigenvalue of the matrix Q , respectively.

Definition 1. The network (6) is said to be globally exponentially stable at x^* , if there are constants $\varepsilon > 0$ and $M \geq 0$ such that for any solution $x(t)$ with the initial point $x(t_0) \in R^n$, one has

$$\|x(t) - x^*\| \leq M \|x(t_0) - x^*\| e^{-\varepsilon(t-t_0)}, \quad \forall t \geq t_0.$$

Definition 2. A mapping $F : R^n \rightarrow R^n$ is said to be monotone on a set Ω if $\forall x, y \in \Omega$

$$(F(x) - F(y))^T(x - y) \geq 0.$$

F is said to be strictly monotone on Ω if the strict inequality above holds whenever $x \neq y$, and strongly monotone on Ω if there exists a constant $\rho > 0$ such that if $\forall x, y \in \Omega$

$$(F(x) - F(y))^T(x - y) \geq \rho \|x - y\|^2.$$

Definition 3. A mapping $F : R^n \rightarrow R^n$ is Lipschitz continuous with constant L if $\forall x, y \in \Omega$

$$\|F(x) - F(y)\| \leq L \|x - y\|.$$

3 Exponential Stability

Now, we can show the existence and unique of solution of the neural network (6).

Lemma 1. The neural network (6) have only one continuous and unique solution with initial point $x(t_0)$.

Proof. Set $F(x) = -x + P_\Omega(-\alpha\nabla f(P_\Omega(x - \alpha\nabla f(x))) + P_\Omega(x - \alpha\nabla f(x)))$.

By the mean-value theorem[16] and the invariance of the norm for the projection operator on closed convex set Ω , we have

$$\begin{aligned}
\|F(x) - F(y)\| &\leq \|y - x\| + \|P_\Omega(-\alpha\nabla f(P_\Omega(x - \alpha\nabla f(x))) + P_\Omega(x - \alpha\nabla f(x))) \\
&\quad - P_\Omega(-\alpha\nabla f(P_\Omega(y - \alpha\nabla f(y))) + P_\Omega(y - \alpha\nabla f(y)))\| \\
&\leq \|y - x\| + \|(I - \sup_{0 \leq t \leq 1} \alpha\nabla^2 f(x + t(x - y)))(P_\Omega(x - \alpha\nabla f(x)) \\
&\quad - P_\Omega(y - \alpha\nabla f(y)))\| \\
&\leq \|y - x\| + \|(I - \alpha Q)^2(x - y)\| \\
&\leq [1 + \sqrt{\lambda_{max}(((I - \alpha Q)^2)^T(I - \alpha Q)^2)}] \|y - x\| \\
&= L \|x - y\|
\end{aligned}$$

where $L = 1 + \lambda_{max}(((I - \alpha Q)^2)^T(I - \alpha Q)^2)$.

Therefore $F(x)$ is Lipschitz continuous in R^n . So, there is a continuous and unique solution $x(t)$ of the neural network (6).

Lemma 2. If $\lambda_{max}(((I - \alpha Q)^2)^T(I - \alpha Q)^2) < 1$, then the neural network (6) has the unique equilibrium point x^* which satisfies

$$x^* = P_\Omega(-\alpha\nabla f(P_\Omega(x^* - \alpha\nabla f(x^*))) + P_\Omega(x^* - \alpha\nabla f(x^*))).$$

Proof. Let $T(x) = P_\Omega(-\alpha\nabla f(P_\Omega(x - \alpha\nabla f(x))) + P_\Omega(x - \alpha\nabla f(x)))$. Then

$$\begin{aligned}
\|T(x) - T(y)\| &\leq \|(-\alpha\nabla f(P_\Omega(x - \alpha\nabla f(x))) + P_\Omega(x - \alpha\nabla f(x))) - \\
&\quad (-\alpha\nabla f(P_\Omega(y - \alpha\nabla f(y))) + P_\Omega(y - \alpha\nabla f(y)))\| \\
&\leq \|(I - \alpha Q)[(P_\Omega(x - \alpha\nabla f(x)) \\
&\quad - P_\Omega(y - \alpha\nabla f(y)))]\| \\
&\leq \sqrt{\lambda_{max}(((I - \alpha Q)^2)^T(I - \alpha Q)^2)} \|x - y\|. \tag{7}
\end{aligned}$$

Therefore, the mapping $T(x)$ is contractive in R^n . Furthermore, $T(x) = x$ has the unique fixed point[16], that is to say, the neural network (6) has the unique equilibrium point x^* which satisfies

$$x^* = T(x^*).$$

Theorem 1. If there exist positive numbers α such that

$$\sqrt{\lambda_{max}(((I - \alpha Q)^2)^T(I - \alpha Q)^2)} < 1,$$

then the neural network (6) is globally exponentially stable to the unique equilibrium point x^* .

Proof. By Lemma 1 and Lemma 2, we know that the neural network (6) has the solution $x(t)$ with initial point $x(t_0)$ and unique equilibrium point x^* .

Let $x(t)$ is any solution of the network (1) with any initial function $x(t_0)$, and x^* denotes the equilibrium point of the neural network (6)

By (6), we get

$$\begin{aligned} x(t) = & e^{-I(t-t_0)}x_0 + \int_{t_0}^t e^{-I(t-s)}[P_\Omega(-\alpha\nabla f(P_\Omega(x(s) - \alpha\nabla f(x(s)))) \\ & + P_\Omega(x(s) - \alpha\nabla f(x(s))))]ds, \end{aligned} \quad (8)$$

and

$$\begin{aligned} x^* = & e^{-I(t-t_0)}x^* + \int_{t_0}^t e^{-I(t-s)}[P_\Omega(-\alpha\nabla f(P_\Omega(x^* - \alpha\nabla f(x^*))) \\ & + P_\Omega(x^* - \alpha\nabla f(x^*)))]ds. \end{aligned} \quad (9)$$

By (8) and (9), we obtain

$$\begin{aligned} x(t) - x^* = & e^{-I(t-t_0)}(x_0 - x^*) + \int_{t_0}^t e^{-I(t-s)}[P_\Omega(-\alpha\nabla f(P_\Omega(x(s) - \alpha\nabla f(x(s)))) \\ & + P_\Omega(x(s) - \alpha\nabla f(x(s)))) - P_\Omega(-\alpha\nabla f(P_\Omega(x^* - \alpha\nabla f(x^*))) \\ & + P_\Omega(x^* - \alpha\nabla f(x^*)))]ds, \end{aligned} \quad (10)$$

Then, by (7), we have

$$\begin{aligned} \|x(t) - x^*\| = & \|e^{-I(t-t_0)}(x_0 - x^*) + \int_{t_0}^t e^{-I(t-s)} \\ & \times [P_\Omega(-\alpha\nabla f(P_\Omega(x(s) - \alpha\nabla f(x(s)))) + P_\Omega(x(s) - \alpha\nabla f(x(s)))) \\ & - P_\Omega(-\alpha\nabla f(P_\Omega(x^* - \alpha\nabla f(x^*))) + P_\Omega(x^* - \alpha\nabla f(x^*)))]ds\| \\ \leq & e^{-I(t-t_0)}\|x_0 - x^*\| + \int_{t_0}^t e^{-I(t-s)} \\ & \times \|P_\Omega(-\alpha\nabla f(P_\Omega(x(s) - \alpha\nabla f(x(s)))) + P_\Omega(x(s) - \alpha\nabla f(x(s)))) \\ & - P_\Omega(-\alpha\nabla f(P_\Omega(x^* - \alpha\nabla f(x^*))) + P_\Omega(x^* - \alpha\nabla f(x^*)))\|ds \\ \leq & e^{-I(t-t_0)}\|x_0 - x^*\| + \int_{t_0}^t e^{-I(t-s)}\sigma(I - \alpha Q)^2\|x(s) - x^*\|ds. \end{aligned} \quad (11)$$

Therefore,

$$\begin{aligned} e^t\|x(t) - x^*\| \leq & e^{t_0}\|x_0 - x^*\| + \sqrt{\lambda_{max}(((I - \alpha Q)^2)^T(I - \alpha Q)^2)} \\ & \times \int_{t_0}^t e^s\|x(s) - x^*\|ds. \end{aligned} \quad (12)$$

Applying Gronwall inequality [] to (12) yields

$$\|x(t) - x^*\| \leq \|x_0 - x^*\| e^{-\varepsilon(t-t_0)} \quad (13)$$

where $\varepsilon = 1 - \sqrt{\lambda_{\max}(((I - \alpha Q)^2)^T(I - \alpha Q)^2)}$ is exponential convergence rate.

It follows from definition 1 that the neural network (6) converges globally exponentially to the unique equilibrium x^* . This completes the proof.

Corollary 1. If $Q > 0$ and $\alpha < 2/\lambda_{\max}(Q)$, then the neural network (6) is globally exponentially stable to the unique equilibrium point x^* .

Proof. It is obvious that if $Q > 0$ and $\alpha < 2/\lambda_{\max}(Q)$, then

$$\lambda_{\max}((I - \alpha Q)^2) < 1.$$

Furthermore, we have

$$\sqrt{\lambda_{\max}(((I - \alpha Q)^2)^T(I - \alpha Q)^2)} < 1.$$

Therefore, by Theorem 1, we get it.

Remark 2. If the conditions of Corollary 1 hold, the LVI (II) is monotone and the equilibrium x^* of the modified neural network (6) is either the solution of the LVI (II) or the optimization solution of the optimization problem (17).

Remark 3. To speed the convergence of the modified neural network (6), we can design the following projection neural network

$$\frac{dx}{dt} = N[-x + P_{\Omega}(-\alpha \nabla f(P_{\Omega}(x - \alpha \nabla f(x))) + P_{\Omega}(x - \alpha \nabla f(x)))] \quad (14)$$

where $N > 0$, $\lambda_{\min}(N) > 1$.

The convergence proof of the neural network (14) is similar to one of Theorem 1 and here omitted.

4 Examples

In this section, two examples will be given to show the validity of our results.

Example 1. Consider the following linear variational inequality (LVI): find $x^* \in \Omega$ such that

$$(Qx^* + c)^T(x - x^*) \geq 0 \quad \forall x \in \Omega \quad (15)$$

where $Q = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$, and $c = (-1 \ -1 \ -1)^T$, and

$$\Omega = \{x \in R^3 \mid 0 \leq x_i \leq 1, i = 1, 2, 3.\} \quad (16)$$

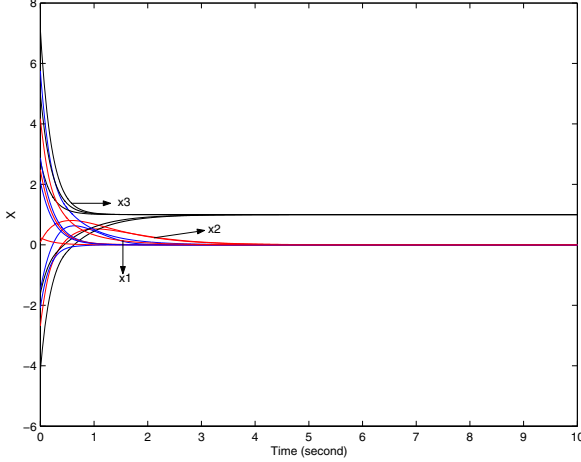


Fig. 1. Dynamical behavior of the neural network (6) with 5 random initial conditions

Let $\alpha = 0.9$, then $\sqrt{\lambda_{\max}(((I - \alpha Q)^2)^T(I - \alpha Q)^2)} = 0.6786 < 1$. Thus, the condition of Theorem 1 holds. The modified projection neural networks (6) and (14) converges globally exponentially to the solution $(0, 0, 1)$ of the linear variational inequality (15). Figure 1 demonstrates the neural network designed converges exponentially to $(x_1^*, x_2^*, x_3^*) = (0, 0, 1)$.

Example 2. Consider the following quadratic programming problem

$$\begin{aligned} \text{minimize} \quad & f(x) = \frac{1}{2}x^T Qx + c^T x \\ \text{subject to} \quad & x \in \Omega \end{aligned} \quad (17)$$

$$Q = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix}, \text{ and } c = (2 \ -1 \ 4)^T, \text{ and}$$

$$\Omega = \{x \in R^3 \mid -2 \leq x_i \leq 2, i = 1, 2, \text{ and } 3 \leq x_3 \leq 5\}. \quad (18)$$

It is obvious that $Q > 0$. Take $\alpha = 0.2 < 2/\lambda_{\max}(Q) = 2/4.4142 = 0.4531$, so the conditions of Corollary 1 hold. According to Corollary 1, the modified projection neural networks (14) with $N = 4I$ converges globally exponentially to the solution $(-2, 0, 3)$ of the optimization problem (17). Figure 2 shows the neural network designed converges exponentially to $(x_1^*, x_2^*, x_3^*) = (-2, 0, 3)$. Compared with the above simulation, we use the neural network (4) to compute the optimization problem in this example. Fig 3 depicts the trajectories $x(t)$ of the neural network (4). From the proof of the Theorem 1 and demonstrate of the Example 2, We can conclude that the convergence speed of the modified projection neural network (6) is faster than one of the projection neural network (4).

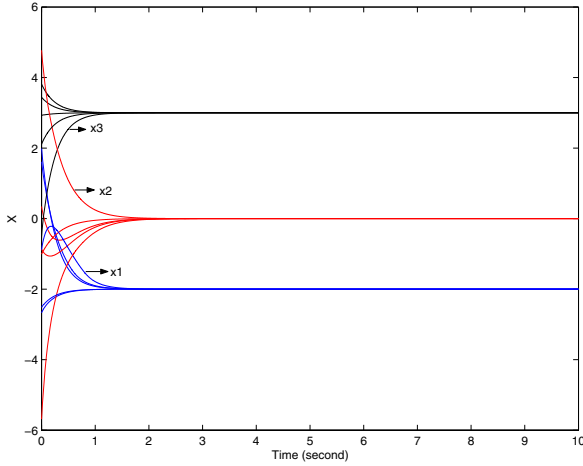


Fig. 2. Transient behavior of the neural network (14) with 5 random initial conditions

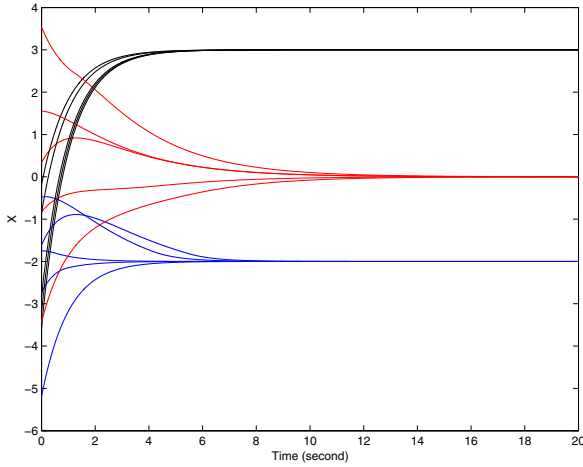


Fig. 3. Trajectories of the neural network (14) with 5 random initial conditions

5 Conclusions

In this paper, the modified neural network (6) are studied. The Theorem 1 and Corollary 1 ensure that the neural network (6) ultimately converge to the solution of variation inequality (1) with the constrained condition (18) and quadratic optimization (6). These results are very easy to verified. Hence, it is very convenience in application. Finally, simulations demonstrate the validity and feasibility of our proposed neural network.

Acknowledgments

The work was supported by the Scientific Innovation TeamProject of Hubei Provincial Department of Education (T200809), the Scientific Research Projects of Hubei Provincial Department of Education (Q200713001) and Hubei Natural Science Foundation (No.2008CDZ046).

References

1. Tank, D.W., Hopfield, J.J.: Simple Neural Optimization Networks: An A/D Converter, Signal Decision Circuit, and a Linear Programming Circuit. *IEEE Trans. Circuits and Systems* 33, 533–541 (1986)
2. Kennedy, M.P., Chua, L.O.: Neural Networks for Nonlinear Programming. *IEEE Trans. Circuits and Systems* 35, 554–562 (1986)
3. Rodriguez-Vazquez, A., Dominguez-Castro, R., Rueda, A., Huertas, J.L., Sanchez-Sinencio, E.: Nonlinear Switched-capacitor Neural Networks for Optimization Problems. *IEEE Trans. Circuits Syst.* 37, 384–397 (1990)
4. Gafni, E.M., Bertsekas, D.P.: Two Metric Projection Methods for Constraints Optimization. *SIAM J. Contr. Optim.* 22, 936–964 (1984)
5. Xia, Y., Wang, J.: Global Exponential Stability of Recurrent Neural Networks for Solving Optimization and Related Problems. *IEEE Trans. Neural Networks* 4, 1017–1022 (2000)
6. Xia, Y., Feng, G.: An Improved Neural Network for Convex Quadratic Optimization with Application to Real-time Beamforming. *Neurocomputing* 64, 359–374 (2005)
7. Xia, Y., Wang, J.: A Recurrent Neural Network for Solving Linear Projection Equations. *Neural Network A* 13, 337–350 (2000)
8. Xia, Y., Leng, H., Wang, J.: A Projection Neural Network and Its Application to Constrained Optimization Problems. *IEEE Trans. Circuits Syst.* 49, 447–458 (2002)
9. Xia, Y., Wang, J.: A General Projection Neural Network for Solving Monotone Variational Inequalities and Related Optimization Problems. *IEEE Trans. Neural Networks* 15, 318–328 (2004)
10. Zhang, S., Constantinides, A.G.: Lagrange Programming Neural Networks. *IEEE Trans. Circuits and Systems II* 39, 441–452 (1992)
11. Chen, Y.H., Fang, S.C.: Neurocomputing with Time Delay Analysis for Solving Convex Quadratic Programming Problems. *IEEE Trans. Neural Networks* 11, 230–240 (2000)
12. Tao, Q., Liu, X., Cui, X.: A Linear Optimization Neural Network for Associative Memory. *Applied Mathematics and Computation* 171, 1119–1128 (2005)
13. Effati, S., Nazemi, A.R.: Neural Networks Models and Its Application for Solving Linear and Quadratic Programming Problems. *Applied Mathematics and Computation* 172, 305–331 (2006)
14. Hu, X., Wang, J.: Design of General Projection Neural Networks for Solving Monotone Linear Variational Inequalities and Linear and Quadratic Optimization Problems. *IEEE Trans. Systems, Man, and Cybernetics-Part B: Cybernetics* 37, 1414–1421 (2007)
15. Bertsekas, D.P.: *Parallel and Distributed Computation: Numerical Methods*. Prentice-Hall, Englewood Cliffs (1989)
16. Halanay, F.A.: *Differential Equations, Stability, Oscillation, Timelags*. Academic Press, NewYork (1996)

Diversity Maintenance Strategy Based on Global Crowding

Qiong Chen, Shengwu Xiong*, and Hongbing Liu

School of Computer Science and Technology,
Wuhan University of Technology, Wuhan 430070, P. R. China
xiongsw@whut.edu.cn

Abstract. In the design of multi-objective evolutionary algorithm, the diversity maintenance is essential to access the convergence of multi-objective optimization solutions. This paper presents a new diversity maintenance strategy based on global crowding, which is addressed for pruning non-dominated solutions as well as preserving a wide-spread distributed solution set and maintaining population diversity. Later on, inspired by the conception of entropy in information theory, the entropy metrics is defined and applied to assess the proposed strategy. Two-dimensional and multi-dimensional numerical experiment results demonstrate that the proposed strategy shows better performance in the entropy reduction and losses of uniform distribution than traditional diversity maintenance strategies.

Keywords: Global Crowding, Diversity Maintenance Strategies, Entropy Metrics, Multi-Objective Evolutionary Algorithm.

1 Introduction

There are a large variety of living things on earth, which is known as biological diversity. Biological diversity is an objective fact. The so-called biological diversity, which includes species diversity, genetic diversity and ecosystem diversity, refers to the diversity of animals, plants, and microorganisms on earth as well as their heredity and mutation. The random variation of gene frequency in small population would give rise to the phenomenon of “genetic drift”. Accurately, it is just like a long-term bottleneck effect. This effect shall repeatedly ruin hetero-zygosity (that is to increase homo-zygosity), reduce the mutation force and eventually result in the loss of genes and allelic genes. This will lead to the loss of genetic diversity. From the loss mechanism of biological diversity, we know that genetic drift is a key factor of causing the loss of genetic diversity, and the degree of drift is often the function of population size. From the perspective of population diversity maintenance, the population diversity is in proportion to the population size. Due to the limitation of computational resource, the population size can not be infinite, it is only be maintained at a reasonable standard. Therefore increasing the population size is not an effective way to solve diversity problems. The multi-objective evolutionary algorithm is to find wide-spread and uniform distributed optimum Pareto front as much as possible [1]. At the same time, in

* Corresponding author

order to avoid the convergence of population to single individual during the evolutionary process, corresponding strategies must be designed to maintain the diversity of population. Researchers have conceived several different strategies through which algorithm can finally find wide-spread and uniform distributed solutions. Among of these researches, NSGA-II [2], SPEA-II [3] and PAES [4] have made significant contributions to multi-objective evolutionary optimization.

This paper proposes a diversity evaluation strategy based on global crowding—Global Crowding Algorithm (GCA). GCA is used for individual's selection process to ensure the wide-spread and uniform distribution of individual and then maintain the diversity of population in the evolutionary process. The crowding measurement of individual is different from the Crowding Distance (CD) in one neighbor in NSGA-II and the nearest neighbor in SPEA-II, the proposed strategy uses the individual global crowding metrics instead of the local one. The individual crowding strength is decided by all other individual rather than a part of individual in the population.

The paper is organized as follows: Related work on traditional diversity maintenance strategies and diversity entropy metrics based on density estimation is described in Section 2. Section 3 proposes a new diversity maintenance strategy based on global crowding. Two-dimensional and multi-dimensional numerical experimental results are analyzed and concluded in Section 4 and Section 5.

2 Classical Density Estimation and Entropy Metrics

2.1 Classical Density Estimation

Generally speaking, in evolutionary algorithm, population would often converge to a single solution in implementing evolutionary process due to some random errors of evolutionary arithmetic operators. To avoid this situation, the diversity maintenance strategy of most evolutionary algorithms in current generation population is to use individual's density information in selection process, which decides the probability of selecting to the next generation. If the individual has high density in its neighbor, then it will have small chance to be selected and replicated to the next generation. This paper summarized two common kinds of density estimation of classical multi-objective evolutionary algorithms, which are the kernel estimation and the nearest neighbor estimation.

The kernel estimation is to define a point's neighborhood scope through a kernel function K which uses the distance to another point as the parameter. In practice, $\sum k(d_i)$ — the distance (d_i) between any individual and another individual i can be calculated by the mapping of the kernel function k . $\sum k(d_i)$ represents the individual density estimation. NSGA-II algorithm [2] is a typical kernel estimation method, in which Deb used the crowding distance to estimate density. The density information is calculated by individual's crowding distance. The Crowding Distance (CD) of the i -th solution in its front is the average side-length of the cuboid.

The nearest neighbor estimation method calculates the distance between the given point and its k -th nearest neighbor and then estimates the density in defined neighbor. The nearest neighbor estimation method was first proposed by B.W.Silverman.

SPEA-II is an adaptation of the k -th nearest neighbor method, where the density at any point is a decreasing function of the distance to the k -th nearest data point. In this paper, this kind of crowding evaluation algorithm is called k -th algorithm for short.

2.2 Entropy Metrics of Diversity

We denote the population set as A , which is divided into Q parts $S = \{S_1, S_2, \dots, S_Q\}$,

$$P_i = \frac{|S_i|}{|A|}. \text{ Then the population diversity entropy is defined as } H = -\sum_{i=1}^Q P_i \lg(P_i). \text{ Pro-}$$

vided the number of population $|A|$ is constant, from the definition of diversity entropy, it can be concluded that the value of P_i is close if the individuals can be uniform distributed in each divided subset [5]. That is to say a high uncertainty of the distribution of individuals means that the population has a high entropy value H . Therefore, in the multi-objective evolutionary algorithms, great entropy value corresponds to well-diversity of population in decision space. For this reason, the entropy value is expected to be great as far as possible to improve the population diversity. In the multi-objective evolutionary process, individuals distributed in population space will gradually approach to the optimal Pareto front. The entropy value will be decreased with the population converging to Pareto optimal front in the evolutionary process. To keep solution uniformly distributed among Pareto optimal front, the entropy value could be decreased slowly in the process of the individuals converging to the optimal solution. Therefore, entropy is a good alternative of metric of population diversity.

3 Diversity Maintenance Strategy Based on Global Crowding

The crowding evaluation function often used in diversity maintenance strategy is generally summarized as density estimation method. The density here can be understood similar to that in physics. The density-like evaluation algorithm only depends on partial property of local individuals but not global individuals. It is very important whether the diversity have global property in the evolutionary process. The degree of an individual's crowding is determined by global influence strength of all other individual in the population rather than local neighbor individuals. Thus, due to the disadvantages of local density-like evaluation, a new global estimation algorithm based on global crowding is proposed to keep the population diversity.

Following definitions and conclusions are given to describe the characteristics of the proposed scheme.

Definition 1. For a population vector set A , if $a_i \in A, i = 1, 2, \dots, |A|$, $|A| = N$, $a_i^{(k)}$ denotes the k -th nearest individual to a_i , and $a_i^{(k)} \in A$, $a_i^{(k)} \neq a_i$, $d_i^{(k)}$ denotes the distance between a_i and the k -th nearest individual, then

$$d_i^{(k)} = |a_i^{(k)} - a_i|, k = 1, 2, \dots, |A| - 1$$

The following conclusion can be easily derived from the above definition:

Lemma 1. For the population vector set A , if

$$a_i \in A, i = 1, 2, \dots, |A|$$

then

$$d_i^k \leq d_i^{(k+1)}, k = 1, 2, \dots, |A| - 2$$

Obviously, in the whole set, if we sort by distances between all the other individuals and a_i in ascending order, $d_i^{(k)}$ (the distance between the k -nearest individual and a_i) is not greater than that of $d_i^{(k+1)}$ (the distance between the $k+1$ -nearest individual and a_i).

Definition 2. For the population vector set A , if $a_i \in A, i = 1, 2, \dots, |A|$, it can be known from definition1 that the k -th nearest individual to a_i is $a_i^{(k)}$, which also corresponds to a certain individual a_j in the population set A . The expression of crowding evaluation value of a_j to a_i is defined as follows:

$$c_{ij} = c_i^{(k)} = g(k)d_i^{(k)}, k = 1, 2, \dots, |A| - 1$$

Wherein, c_{ij} denotes the influencing strength of a_i by the arbitrary individual $a_j \in A$. $d_i^{(k)}$ (the distance between a_i and a_j) denotes the k -th distance ascending order among all other individuals' distance to a_i .

The function $g(k)$ given in Definition 2 is a decreasing function. The specific meaning of function $g(k)$ has direct influence on evolutionary algorithm's ultimate result. Generally, the influencing strength of other individuals on considering individual is directly related to their distance. The distance between a_i and a_j is represented as d_{ij} . Taking into account that the influence value of evaluation function is small if d_{ij} is large, following two rules are derived.

Rule 1. For $\forall j$ and , if $d_{ij} < d_{ik}$, then $E(c_{ij}) > E(c_{ik})$. For individual a_i , the individual a_j that has shorter distance to a_i must have larger influence value than the individual a_k that has longer distance to a_i .

Rule 2. For $\forall j, k$, if $d_{ij} < d_{ik}$, then

$$E(c_{ij}) > E\left(\sum_{l=k}^N c_{il}\right)$$

Rule 3. It makes clear that mathematical expectation of the influence value of the k -th furthest individual to a_i is larger than sum of the mathematical expectation's of the influence value of many individuals which are further than the k -th individual. From Rule 1 and Lemma 1 we can conclude that $g(k)$ is obviously a monotonic decreasing function.

The table 1 below lists several recommended functions $g(k)$ and their characteristics.

Table 1. Common laws of function $g < 0$

Function	Rule1	Rule2
$g_1(k) = \frac{1}{k}$	If $M \geq 2$, satisfied	Not satisfied
$g_2(k) = \frac{1}{2^k}$	Satisfied	Not satisfied
$g_3(k) = \frac{1}{k^2}$	Satisfied	Not satisfied
$g_4(k) = \frac{1}{k!}$	Satisfied	If $k \geq 2$, satisfied

From table 1, we know that if $k \geq 2, M \geq 2$, then the function $g(k) = 1/k!$ meets both of the two rules. In this paper, the function $g(k) = 1/k!$ is selected as the main function in the proposed crowding evaluation algorithm. The distance order relation is the key factor to the proposed algorithm (c_{ik}). The influencing strength of a_i in Definition 2 corresponds to the certain c_{ij} , so the following inferences can be derived :

Theorem 1. $c_i = \sum_{i \neq j}^N c_{ij} = \sum_{i \neq j}^{N-1} c_i^{(k)}$

The above theorem finally gives the crowding evaluation function of global crowding algorithm. The crowding evaluation function of global crowding algorithm is specific defined as follows:

Definition 3. For the individual a_i in the population vector set A , its global crowding evaluation is defined as:

$$C_i = \sum_{k=1}^{N-1} g(k)d_i^{(k)}$$

C_i is the sum of the influence value by the all other individuals $a_j \in A$, $j = 1, 2, \dots, N, j \neq i$.

For any individual $a_i \in A$, where A represents the population vector set including N individuals, the proposed algorithm firstly calculates the distance from a_i to every other individual, then sorts the $N-1$ distance values by ascending order, finally calculates the corresponding global crowding value c_i . The global crowding algorithm is detailed as follows:

```

Procedure Global Crowding (Input: A) Output: C;
size=|A|;
for i = 1 to size
  for j = 1 to size
    D[i][j] = |A[i] -A[j]|;
  end for j;

```



```

Sort(D[i]);
t = 0;
for j = 2 to size
k = j - 1;
t = t + g(k)*D[i][j];
end for j;
C[i] = t ;
end for i;
end Procedure

```

4 Numerical Experimentation

In the process of constructing multi-objective evolutionary algorithm, most of algorithms use the strategy of randomly initializing population. But some random factors may affect the population diversity, this paper introduces the method of diversity based on the strategy of randomly initializing population. The diversity method makes the population maintain a uniform distributed initial population and speed up the convergence. In our numerical experiments, the function $g(x)$ is chosen as $g(x) = \frac{1}{2^x}$, $g(x) = \frac{1}{k^x}$, $g(x) = \frac{1}{k!}$ (hereinafter referred to as Global-2K, Global -K2,

Global-E). The performance was compared among the proposed Global Crowding Algorithm, traditional CD and k -th Algorithm in two dimensional and multi-dimensional numerical experiments. Here we choose the statistical results of entropy loss (ΔH) and the amount of loss (λ) of uniform distribution as the comparison index.

In M -dimensional space, the variable bound in each dimension is restricted to $[0,1]$. N points are randomly generated, conforming to the specified distribution. Each variable is divided into $\lceil \sqrt[M]{N} \rceil$ even part in each dimension. Thus about N hypercube subspaces are constructed (Generally, N is M -th power of an integer). Then $N/2$ individuals are reserved by specified diversity maintenance strategy.

4.1 Two-Dimensional Numerical Experiments

Numerical experiments results are compared among with the crowding evaluation algorithms CD, k -th, Global-2K, Global-K2, Global-E in two-dimensional space. The figure 1 to figure 5 show comparison results among different diversity maintenance strategies, where entropy value is set to 4.0158 and keeps 50 of 100 individuals. The parameters are set to $M = 2, N = 100$. All results are statistical results in 1000 times running. The uniform distribution is selected to randomly initialize the population. In the following figure, filled circles denote reserved individuals and empty circles denote discarded individuals. Statistical comparison results of entropy loss (ΔH) and the amount of loss (λ) of uniform distribution are listed in table 2 and table 3.

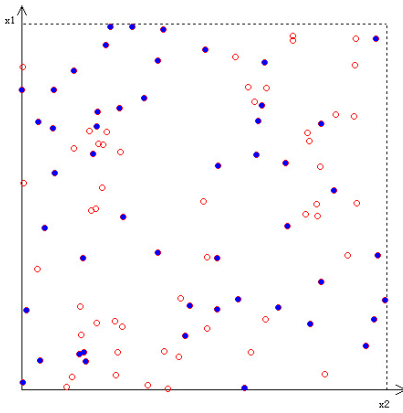
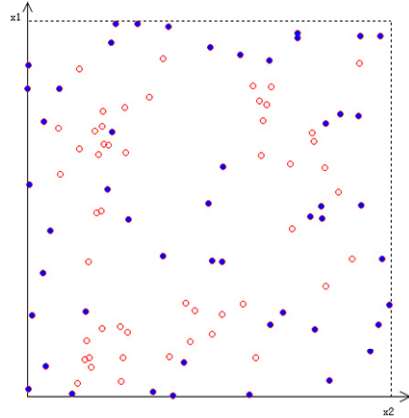
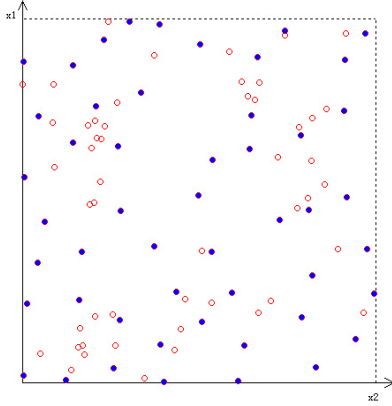
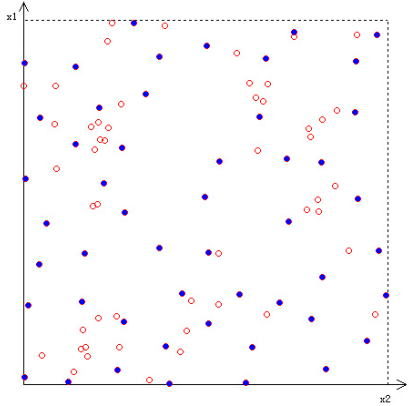
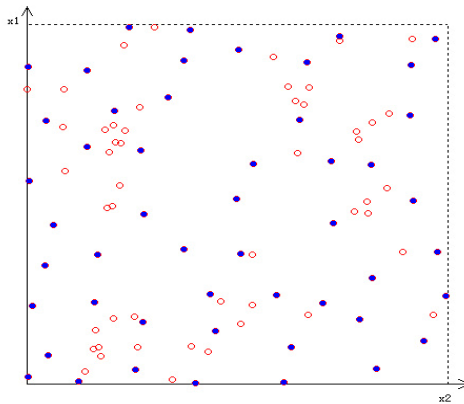
**Fig. 1.** CD ($H' : 3.5965, \lambda : 0.31$)**Fig. 2.** K-th ($H' : 3.5688, \lambda : 0.36$)**Fig. 3.** Global -2K ($H' : 3.8843, \lambda : 0.16$)**Fig. 4.** Global -K2 ($H' : 3.8843, \lambda : 0.16$)**Fig. 5.** Global -E ($H' : 3.9120, \lambda : 0.16$)

Table 2. Entropy Loss of Uniform Distribution (ΔH)

	min	max	mean	D
CD	0.20881	0.62906	0.42974	0.06850
K-th	0.27542	0.80657	0.51613	0.07544
Global -2K	0.02038	0.33950	0.17301	0.05466
Global -K2	0.02038	0.34290	0.17606	0.05525
Global -E	-0.00730	0.32904	0.16372	0.05361

Table 3. Loss of Uniform Distribution (λ)

	min	max	mean	D
CD	0.42	0.19	0.3021	0.03670
Kth	0.46	0.29	0.3767	0.03276
Global -2K	0.33	0.10	0.2201	0.04102
Global -K2	0.34	0.10	0.2213	0.04113
Global -E	0.32	0.11	0.2144	0.04126

Table 4. Entropy Reduction of Uniform Distribution (ΔH)

	min	max	mean	D
CD	0.3119	0.5310	0.4089	0.04406
K-th	0.4101	0.6738	0.5270	0.05248
Global -2K	0.08896	0.3135	0.2109	0.04190
Global -K2	0.05844	0.3303	0.2109	0.04498
Global -E	0.04561	0.3135	0.2094	0.04261

Table 5. Losses of Uniform Distribution (λ)

	min	max	mean	D
CD	0.3565	0.2362	0.2994	0.02568
K-th	0.4444	0.3241	0.3805	0.02428
Global -2K	0.2870	0.1481	0.2267	0.02932
Global -K2	0.3102	0.1481	0.2298	0.03122
Global -E	0.2963	0.1435	0.2264	0.02994

4.2 Multi-dimensional Numerical Experiments

All numerical experiments run on the same as the case in two-dimensional experiments except $M=3$, $N=300$. Statistical comparison results of entropy loss (ΔH) and the amount of loss (λ) of uniform distribution are listed in table 4 and table 5.

5 Conclusion

Numerical experiments demonstrate that the proposed crowding evaluation algorithm can achieve better results in diversity maintenance. It can be derived from the method of entropy metrics that greater entropy value corresponds to better diversity after implementing diversity maintenance strategy. Two-dimensional and multi-dimensional numerical experiment results demonstrate that the proposed strategy shows better performance in entropy reduction and losses of uniform distribution than traditional diversity maintenance strategies. Our further numerical experimental results on multi-objective evolutionary optimization benchmark testing functions which are ZDT1, ZDT2, ZDT3, ZDT4 and ZDT6 show that the proposed diversity maintenance strategy also demonstrates superior performance in the convergence of solutions and the homogeneity of distribution of non-dominated solutions.

The proposed scheme achieves satisfied performance both in convergence and diversity maintenance. But in comparison with the density-based method, global crowding evaluation strategy is computational cost. This also proves the "No Free Lunch" theorem. The future work will concentrate on the improving of the computational efficiency and achieving the trade-off between the diversity and convergence.

Acknowledgments. This work was supported in part by National Science Foundation of China (Grant No.40701153) and Wuhan International Cooperation and Communication Project (Grant No.200770834318).

References

1. Pareto, V.: *Cours d'économie politique*. Rouge, Lausanne, Switzerland (1896)
2. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A Fast and Elitist Multi-objective Genetic Algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation* 6, 182–197 (2002)
3. Zitzler, E., Laumanns, M., Thiele, L.: SPEA2: Improving the Strength Pareto Evolutionary Algorithm for Multi-objective Optimization. In: *Proceeding of Evolutionary Methods for Design, Optimisation and Control with Applications to Industrial Problems*, pp. 95–100 (2001)
4. Knowles, J., Corne, D.: The Pareto Archived Evolution Strategy: A New Baseline Algorithm for Pareto Multi-objective Optimization. In: *Proceeding of the Congress on Evolutionary Computation*, Piscataway, New Jersey, pp. 98–105. IEEE Press, Los Alamitos (1999)
5. Foster, I., Kesselman, C., Nick, J., Tuecke, S.: *The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration*. Technical report, Global Grid Forum (2002)
6. Farhang-Mehr, A., Azarm, S.: Diversity Assessment of Pareto Optimal Solution Sets: An Entropy Approach. In: *Proceedings of the Congress on Evolutionary Computation*, pp. 723–728 (2002)

Hybrid Learning Enhancement of RBF Network Based on Particle Swarm Optimization

Sultan Noman Qasem and Siti Mariyam Shamsuddin

Soft Computing Group, Faculty of Computer Science and Information System,
UTM, Malaysia
sultannoman@yahoo.com

Abstract. This study proposes RBF Network hybrid learning with Particle Swarm Optimization for better convergence, error rates and classification results. In conventional RBF Network structure, different layers perform different tasks. Hence, it is useful to split the optimization process of hidden layer and output layer of the network accordingly. RBF Network hybrid learning involves two phases. The first phase is a structure identification, in which unsupervised learning is exploited to determine the RBF centers and widths. The second phase is parameters estimation, in which supervised learning is implemented to establish the connections of weights between the hidden layer and the output layer. The incorporation of PSO in RBF Network hybrid learning is accomplished by optimizing the centers, the widths and the weights of RBF Network. The results for training, testing and validation on dataset illustrate the effectiveness of PSO in enhancing RBF Network learning.

Keywords: RBF Network, Hybrid learning, Particle Swarm Optimization.

1 Introduction

Radial Basis Function (RBF) forms a class of Artificial Neural Networks (ANNs), which has certain advantages over other types of ANNs, such as better approximation capabilities, simpler network structures and faster learning algorithms. RBF Network is a fully connected three layer feed forward network, and establishes nonlinearity in the hidden layer neurons. The output layer has no nonlinearity and the connections of the output layer are only weights [1].

Due to their better approximation capabilities, simpler network structures and faster learning algorithms, RBF Networks have been widely applied in many science and engineering fields. It is three layers feedback network, where each hidden unit implements a radial activation function and each output unit implements a weighted sum of hidden units' outputs. Its training procedure is usually divided into two stages. The first stage includes determination of centers and widths of the hidden layer by clustering algorithms such as K-means, vector quantization, decision trees, and self-organizing feature maps. The second stage involves weights establishment by connecting the hidden layer with the output layer. This is determined by Singular Value Decomposition (SVD) or Least Mean Squares (LMS) algorithms [2]. The problem of selecting the appropriate number of basis functions remains a critical issue for RBF

Networks. The number of basis functions controls the complexity and the generalization ability of RBF Networks. RBF Networks with too few basis functions cannot fit the training data adequately due to limited flexibility. On the other hand, those with too many basis functions yield poor generalization abilities since they are too flexible and fit the noise in the training data. The methods mentioned above require designers to fix the structure of networks in advance according to prior knowledge. However it is difficult for designers to achieve optimal architecture.

Clustering algorithms have been successfully used in training RBF Networks such as Optimal Partition Algorithm (OPA) to determine the centers and widths of RBFs. In most traditional algorithms, such as K-means, the number of cluster centers need to be predetermined; hence restricts the real applications of this algorithm. In addition, Genetic Algorithm (GA), Particle Swarm Optimization (PSO) and Self-Organizing Maps (SOM) have also been considered in clustering process [4]. Training technique can be formulated as an optimization problem, which includes the network structure into a set of variables that are used to minimize the prediction error. PSO possess similar attractive features of genetic algorithms such as independence from gradient information of the objective function, the ability to solve complex nonlinear high dimensional problems. Furthermore, they can achieve faster convergence speed and require fewer parameters to be adjusted.

In this paper, PSO algorithm is adapted to auto-configure the structure of RBF Network and obtain the model parameters according to given input-output examples and is explored to enhance RBF Network learning. The paper is structured as followed. In section 2, related work about RBF Network training is introduced. Section 3 presents RBF Network model and parameter selection problem. Section 4 describes PSO algorithm. In Section 5, a proposed PSO-RBF Network model is given. Sections 6 and Section 7 give the experiments setup, results and validation of the proposed model on various datasets. Finally, the conclusion is given in Section 8.

2 Related Work

Although there are many studies in RBF Network training, but research on training of RBF Network with PSO is still fresh. This section presents some existing work of training RBF Network based on Evolutionary Algorithms (EAs) such as PSO especially based on unsupervised learning only (Clustering).

In [10], they have proposed a PSO learning algorithm to automate the design of RBF Networks, to solve pattern classification problems. Thus, PSO-RBF finds the size of the network and the parameters that configure each neuron: center and width of its basis function. Supervised mean subtractive clustering algorithm has been proposed [3] to evolve RBF Networks and the evolved RBF acts as fitness evaluation function of PSO algorithm for feature selection. The method performs feature selection and RBF training simultaneously. PSO algorithm has been introduced [2] to train RBF Network related to automatic configuration of network architecture related to centers of RBF. Two training algorithm were compared. One was PSO algorithm. The other was newrb routine that was included in Matlab neural networks toolbox as standard training algorithm for RBF network.

A hybrid PSO (HPSO) was proposed [11] with simulated annealing and Chaos search technique to train RBF Network. The HPSO algorithm combined the strong ability of PSO, SA, and Chaos. An innovative Hybrid Recursive Particle Swarm Optimization (HRPSO) learning algorithm with normalized fuzzy c-mean (NFCM) clustering, PSO and Recursive Least Squares (RLS) has been presented [13] to generate RBF networks modeling system with small numbers of descriptive RBFs for fast approximating two complex and nonlinear functions. On the other hand, a newly evolutionary search technique called Quantum-Behaved Particle Swarm Optimization, in training RBF Network has been used [12]. The proposed QPSO-Trained RBF Network was test on nonlinear system identification problem.

Unlike previous studies, this research shares consideration of parameters of RBF (unsupervised learning) which are centers and length of width or spread of RBFs with different algorithms such as K-means and K-nearest neighbors or standard deviations algorithms respectively. However, training of RBF Network need to be enhanced with PSO to optimize the centers and widths values which are obtained from the clustering algorithms and PSO also used to optimize the weights which connect between hidden layer and output layer (supervised learning). Also this paper has been presented to train, test and validate the PSO-RBF Network on the datasets.

3 Architecture of RBF Network

RBF Networks were introduced into the neural network byBroomhead [14]. Due to the better approximation capabilities, simpler network structures and faster learning algorithms, RBF Networks have been widely used in many fields.

A RBF Network has three-layer architecture. The input layer which consists of a set of source nodes connects the network to the environment. The hidden layer consists of hidden neurons (radial basis units), with radial activation functions. The activation of a hidden neuron is determined by computing the distance (usually by using the Euclidean norm) between its center vector and the vectors which are yielded by the activation of the input layer. The activation of a neuron in the output layer is determined by computing the weighted sum of outputs of hidden layer. The RBF Network form with linear combination of Gaussian functions is shown in the following.

$$o_i(x) = \sum_{k=1}^N w_{ik} \exp\left\{-\frac{\|x - c_k\|^2}{2\sigma_k^2}\right\}, i = 1, 2, \dots, m \quad (1)$$

Where $\|\dots\|$ represents Euclidean norm, c_k , σ_k and w_{ik} are the center, the width of the k -th neuron in the hidden layer and the weights in the output layer respectively, K is the number of neurons in the hidden layer.

RBF Networks are universe function approximators if the centers and widths are set appropriately. The number of radial basis functions affects the performance of RBF Network, networks with too many hidden units overfitting training data and having poor predictive ability while ones with insufficient hidden units having poor approximation power. If the width is narrower than necessary, this will lead to overfitting, on the other hand, a width wider than necessary can result in poor approximation ability and give even worst results.

4 Particle Swarm Optimization (PSO)

PSO algorithm originally introduced by Kennedy and Eberhart in 1995 [5], simulates the knowledge evolution of a social organism, in which each individual is treated as an infinitesimal particle in the n-dimensional space, with the position vector and velocity vector of particle i being represented as $X_i(t) = (X_{i1}(t), X_{i2}(t), \dots, X_{in}(t))$ and $V_i(t) = (V_{i1}(t), V_{i2}(t), \dots, V_{in}(t))$. The particles move according to the following equations:

$$V_{id}(t+1) = W \times V_{id}(t) + c_1 r_1 (P_{id}(t) - X_{id}(t)) + c_2 r_2 (P_{gd}(t) - X_{id}(t)) \quad (2)$$

$$X_{id}(t+1) = X_{id}(t) + V_{id}(t+1) \quad (3)$$

$$i = 1, 2, \dots, M; d = 1, 2, \dots, n$$

Where c_1 and c_2 are the acceleration coefficients, Vector $P_i = (P_{i1}, P_{i2}, \dots, P_{in})$ is the best previous position (the position giving the best fitness value) of particle i known as the personal best position (pbest); Vector $P_g = (P_{g1}, P_{g2}, \dots, P_{gn})$ is the position of the best particle among all the particles in the population and is known as the global best position (gbest). The parameters r_1 and r_2 are two random numbers distributed uniformly in (0, 1). Generally, the value of V_{id} is restricted in the interval $[-V_{max}, V_{max}]$. Inertia weight w was first introduced by Shi and Eberhart in order to accelerate the convergence speed of the algorithm [6].

5 PSO-RBF Network

PSO has been applied to improve RBF Network in various aspects such as network connections, network architecture and learning algorithm. Every single solution of

```

For each particle do
    Initialize particle position and velocity
End for
While stopping criteria are not fulfilled do
    For each particle do
        Calculate fitness value (MSE in RBF Network)
        If fitness value is better than best fitness value pBest in particle
            history then
                Set current position as pBest
        End if
    End for
    Choose as gBest the particle with best fitness value among all particles in
    current iteration
    For each particle do
        Calculate particle velocity based on eq. (2)
        Update particle position(center, width and weight) based on eq. (3)
    End for
End while

```

Fig. 1. PSO-RBF Network Algorithm

PSO called a particle flies over the solution space in search for the optimal solution. The particles are evaluated using a fitness function to seek the optimal solution. Particles values of RBF parameters are then initialized with values which are obtained from the k-means algorithm while particles values of the weights and bias are initialized randomly or from LMS algorithm. The particles are updated accordingly using the equation (2) and (3). The procedure for implementing PSO global version (gbest) is shown in Figure 1. Optimization of RBF network parameters with PSO, the fitness value of each particle is the value of the error function evaluated at the current position of the particle and position vector of the particle corresponds to the parameters matrix of the network.

6 Experiments

6.1 Experimental Setup

The experiments of this work include the standard PSO and BP for RBF Network training. For evaluating all of these algorithms we used five benchmark classification problems obtained from the machine learning repository [9].

Table 1. Execution parameters for PSO

Parameter	Value
Population Size	20
Iterations	10000
W	[0.9,0.4]
C_1	2.0
C_2	2.0

The parameters of the PSO algorithm were set as: weight w decreasing linearly between 0.9 and 0.4, learning rate $c_1 = c_2 = 2$ for all cases. The population size used by PSO was constant. Values selected for parameters are shown in table 1. The parameters of the experiments are described in Table 2.

Table 2. Parameters of the experiments

Parameter	Dataset			
	Balloon	Cancer	Iris	Ionosphere
Train data	12	349	120	251
Test data	4	175	30	100
Validation data	16	175	150	351
Input dimension	4	9	4	34
Output neuron	1	1	3	1
Network Structure	4-2-1	9-2-1	4-3-3	34-2-1

The number of maximum iterations is set differently to bound the number of forward propagations to 4×10^4 and for comparison purposed. The maximum iterations in BP-RBFN is set to 2×10^4 (number of forward propagations = $2 \times$ maximum number of iterations), while the maximum number of iterations in PSO-RBFN is set to 10000 (number of forward propagations = swarm size \times maximum number of iterations) [8]. The stopping criteria are the maximum number of iterations that the algorithm has been reached or the minimum error.

6.2 Experimental Results

This section presents the results of the study on PSO-trained RBF Network and BP-trained RBF Network. The experiments are conducted by using four datasets.

6.2.1 Balloon Dataset

This data is used in cognitive psychology experiment. There are four data sets representing different conditions of an experiment. It contains 4 attributes and 16 instances. The stopping conditions of PSO-RBFN are set to a minimum error of 0.005 or maximum iteration of 10000. Conversely, the stopping conditions for BP-RBFN are set to the minimum error of 0.005 or the iterations have reached to 20000.

Table 3. Result of BP-RBFN and PSO-RBFN on Balloon dataset

	BP-RBFN		PSO-RBFN	
	Train	Test	Train	Test
Learning Iteration	20000	1	3161	1
Error Convergence	0.01212	0.23767	0.0049934	0.16599
Classification (%)	91.27	75.41	95.05	78.95

From Table 3, we conclude that PSO-RBFN converges faster compared to BP-RBFN for the whole learning process. However, both algorithms have converged to the given minimum error. For the classification, it shows that PSO-RBFN is better than BP-RBFN. Figure 2 illustrates the learning process for both algorithms.

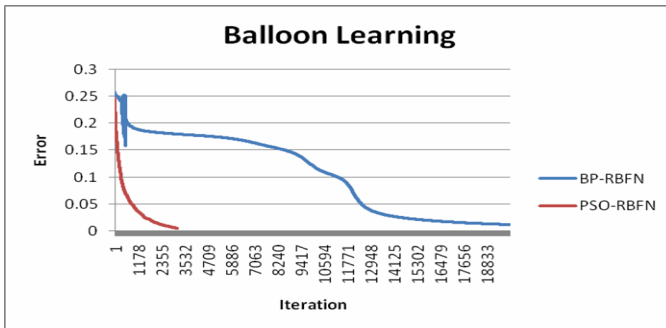


Fig. 2. Convergence of Balloon dataset

6.2.2 Cancer Dataset

The purpose of the breast cancer data set is to classify a tumour as either benign or malignant based on cell descriptions gathered by microscopic examination. It contains 9 attributes and 699 examples of which 485 are benign examples and 241 are malignant examples. The ending conditions of PSO-RBFN are set to minimum error of 0.005 or maximum iteration of 10000. Alternatively, the stopping conditions for BP-RBFN are set to a minimum error of 0.005 or maximum iteration of 20000 has been achieved.

Table 4. Result of BP-RBFN and PSO-RBFN on Cancer dataset

	BP-RBFN		PSO-RBFN	
	Train	Test	Train	Test
Learning Iteration	20000	1	10000	1
Error Convergence	0.03417	0.27333	0.0181167	0.27464
Classification (%)	92.80	70.37	97.65	71.77

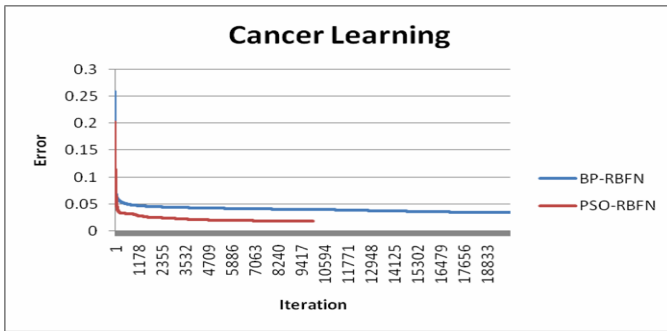


Fig. 3. Convergence of Cancer dataset

In Cancer learning process, from Table 4 shows PSO-RBFN takes 10000 iterations compared to 20000 iterations in BP-RBFN to converge. PSO-RBFN managed to converge at iteration 10000, while BP-RBFN converges at a maximum iteration of 20000 and illustrates that PSO-RBFN is better than BP-RBFN. Figure 3 shows PSO-RBFN significantly reduce the error with small number of iterations.

6.2.3 Iris Dataset

The Iris dataset is used for classifying all the information into three classes which are iris setosa, iris versicolor, and iris virginica. The classification is based on its four input patterns which are sepal length, sepal width, petal length and petal width. Each class refers to a type of iris plant containing 50 instances. For Iris dataset, the minimum error of PSO-RBFN is set to 0.05 or maximum iteration of 10000. While, the minimum error for BP-RBFN is set to 0.05 or the network has reached maximum iteration of 20000. Table 5 shows that BP-RBFN is better than PSO-RBFN.

For Iris learning, both algorithms converge using the maximum number of pre-specified iteration. PSO-RBFN takes 3774 iterations to converge at a minimum error of 0.0499949 while minimum error for BP-RBFN is 0.05000 with 10162 iterations. Figure 4 shows that PSO-RBFN reduces the error with minimum iterations.

Table 5. Result of BP-RBFN and PSO-RBFN on Iris dataset

	BP-RBFN		PSO-RBFN	
	Train	Test	Train	Test
Learning Iteration	10162	1	3774	1
Error Convergence	0.05000	0.04205	0.0499949	0.03999
Classification (%)	95.66	95.78	95.48	95.64

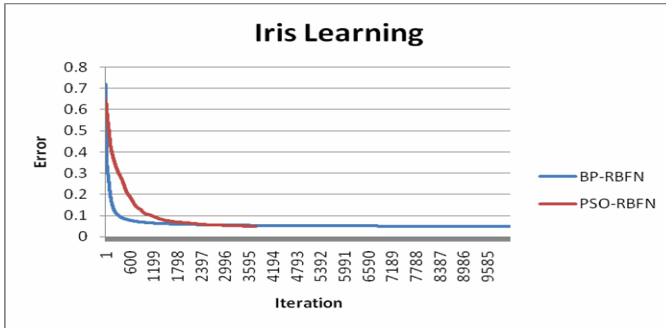


Fig. 4. Convergence of Iris dataset

6.2.4 Ionosphere Dataset

This radar data was collected by a system in Goose Bay, Labrador. This system consists of a phased array of 16 high-frequency antennas with a total transmitted power on the order of 6.4 kilowatts. The targets were free electrons in the ionosphere. "Good" radar returns are those showing evidence of some type of structure in the ionosphere. "Bad" returns are those that do not; For Ionosphere problems, the stopping conditions for BP-RBFN is minimum error of 0.05 or maximum iteration of 20000. The minimum error of PSO-RBFN is 0.05 or maximum iteration of 10000. The experimental results for PSO-based RBFN and BP-based RBFN are shown in Table 6 and Figure 5.

From Ionosphere learning, Table 6 shows PSO-RBFN takes 5888 iterations compared to 20000 iterations in BP-RBFN to converge. In this experiment, PSO-RBFN manages to converge using minimum error at iteration of 5888, while BP-RBFN trapped at the local minima and converges at a maximum iteration of 20000. For the correct classification percentage, it shows that PSO-RBFN result is better than BP-RBFN. Figure 5 shows PSO-RBFN significantly reduce the error with small number

of iterations compared to BP-RBFN. The results for this data are not promising for BP-RBFN since it depends on the data and repeatedly traps in local minima. The local minima problem in BP-RBFN algorithm is usually caused by disharmony adjustments between centers and weights of RBF Network. To solve this problem, the error function has been modified as suggested [7].

Table 6. Result of BP-RBFN and PSO-RBFN on Ionosphere dataset

	BP-RBFN		PSO-RBFN	
	Train	Test	Train	Test
Learning Iteration	20000	1	5888	1
Error Convergence	0.18884	0.23633	0.0499999	0.01592
Classification (%)	62.27	62.71	87.24	90.70

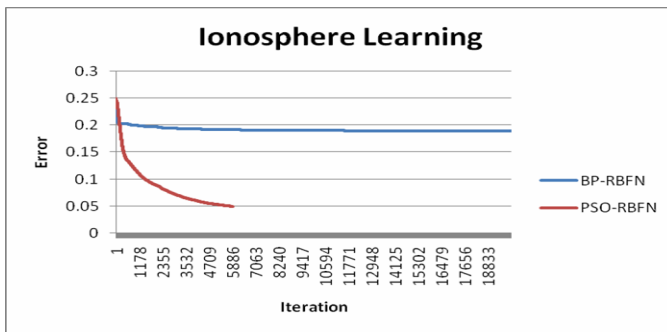


Fig.5. Convergence of Ionosphere dataset

7 Validation Results

In Artificial Neural Network (ANN) methodology, data samples are divided into three sets; training, validation and testing in order to obtain a network which is capable of generalizing and performing well with new cases. There is no precise rule on the optimum size of the three sets of data, although authors agree that the training set

Table 7. Validation Result of BP-RBFN and PSO-RBFN on all dataset

Dataset	BP-RBFN		PSO-RBFN	
	Train	Test	Train	Test
Balloon	0.06004	0.33155	0.00499450	0.27348
Cancer	0.03046	0.04233	0.00541208	0.02733
Iris	0.05000	0.06227	0.0499760	0.05792
Ionosphere	0.20743	0.22588	0.0499953	0.06325

must be the largest. Validations are motivated by two fundamental problems either in model selection or in performance estimation. These results for BP-RBFN and PSO-RBFN on all dataset are shown in Table 7.

8 Conclusion

This paper proposes PSO based Hybrid Learning of RBF Network to optimize the parameters of network. The proposed algorithm is successfully applied on well known dataset. The results obtained are compared with the results of BP-RBFN. It is clear that PSO-RBFN is better than BP-RBFN in term of convergence and error rates. Furthermore PSO-RBFN reached optimum because it reduces the error with minimum iteration and obtains the optimal parameters of RBF Network.

Acknowledgment

This work is supported by Ministry of Higher Education (MOHE) under Fundamental Research Grant Scheme (Vote No. 78243). Authors would like to thank Research Management Centre (RMC) Universiti Teknologi Malaysia, for the research activities and Soft Computing Research Group (SCRG) for the support and incisive comments in making this study a success.

References

1. Leonard, J.A., Kramer, M.A.: Radial Basis Function Networks for Classifying Process Faults. *Control Systems Magazine* 11(3), 31–38 (1991)
2. Liu, Y., Zheng, Q., Shi, Z., Chen, J.: Training Radial Basis Function Networks with Particle Swarms. In: Yin, F.-L., Wang, J., Guo, C. (eds.) *ISNN 2004*. LNCS, vol. 3173, pp. 317–322. Springer, Heidelberg (2004)
3. Chen, J.Y., Qin, Z.: Training RBF Neural Networks with PSO and Improved Subtractive Clustering Algorithms. In: King, I., Wang, J., Chan, L.-W., Wang, D. (eds.) *ICONIP 2006*. LNCS, vol. 4233, pp. 1148–1155. Springer, Heidelberg (2006)
4. Cui, X.H., Polok, T.E.: Document Clustering Using Particle Swarm Optimization. In: *Proceedings of Swarm Intelligence Symposium, 2005*. SIS 2005, pp. 185–191. IEEE, Los Alamitos (2005)
5. Kennedy, J., Eberhart, R.C.: Particle Swarm Optimization. In: *Proceedings of IEEE International Conference on Neural Networks, IV*, Piscataway, NJ, pp. 1942–1948 (1995)
6. Shi, Y., Eberhart, R.C.: A Modified Particle Swarm. In: *Proceedings of 1998 IEEE International Conference on Evolutionary Computation*, Piscataway, NJ, pp. 1945–1950 (1998)
7. Wang, X.G., Tang, Z., Tamura, H., Ishii, M.: A Modified Error Function for the Backpropagation Algorithm. *Neurocomputing* 57, 477–488 (2004)
8. Al-kazemi, B., Mohan, C.K.: Training Feedforward Neural Network Using Multiphase Particle Swarm Optimization. In: *Proceeding of the 9th International Conference on Neural Information Processing 5*, pp. 2615–2619 (2002)
9. Blake, C., Merz, C.J.: *UCI Repository of Machine Learning Databases* (1998), <http://www.ics.uci.edu/~mllearn/MLRepository.html>

10. Qin, Z., Chen, J., Liu, Y., Lu, J.: Evolving RBF Neural Networks for Pattern Classification. In: Hao, Y., Liu, J., Wang, Y.-P., Cheung, Y.-m., Yin, H., Jiao, L., Ma, J., Jiao, Y.-C. (eds.) CIS 2005. LNCS, vol. 3801, pp. 957–964. Springer, Heidelberg (2005)
11. Gao, H., Feng, B., Hou, Y., Zhu, L.: Training RBF Neural Network with Hybrid Particle Swarm Optimization. In: Wang, J., Yi, Z., Żurada, J.M., Lu, B.-L., Yin, H. (eds.) ISNN 2006. LNCS, vol. 3971, pp. 577–583. Springer, Heidelberg (2006)
12. Sun, J., Xu, W., Liu, J.: Training RBF Neural Network via Quantum-Behaved Particle Swarm Optimization. In: King, I., Wang, J., Chan, L.-W., Wang, D. (eds.) ICONIP 2006. LNCS, vol. 4233, pp. 1156–1163. Springer, Heidelberg (2006)
13. Chen, C.Y., Feng, H.M., Ye, F.: Hybrid Recursive Particle Swarm Optimization Learning Algorithm in the Design of Radial Basis Function Networks. *Journal of Marine Science and Technology* 15, 31–40 (2007)
14. Broomhead, D.S., Lowe, D.: Multivariable Functional Interpolation and Adaptive Networks. *Complex System* 2, 321–355 (1988)

Chaos Cultural Particle Swarm Optimization and Its Application

Ying Wang, Jianzhong Zhou^{*}, Youlin Lu, Hui Qin, and Yongchuan Zhang

School of Hydropower and Information Engineering
Huazhong University of Science and Technology
Wuhan, Hubei 430074, China
jzhhbjz111@gmail.com, jz.zhou@hust.edu.cn

Abstract. A new version of the classical particle swarm optimization (PSO), namely, Chaos culture particle swarm optimization (CCPSO), is proposed to overcome the shortcoming of the premature of the classical PSO. The proposed algorithm integrates PSO with the framework of cultural algorithm model. PSO is utilized as the evolution method of population space. Meanwhile, the chaotic search operator is imported to build the knowledge structure of belief space, with which guiding the evolution process of the proposed algorithm, moving particles to the global optimal solution can be more effective. Then, the proposed algorithm is tested with typical test functions. The result shows that the global searching ability of CCPSO is better than that of PSO. In the last part of the paper, CCPSO was applied to the optimal operation of cascade hydropower station. The operation result shows the feasibility and high efficiency of the proposed algorithm, while compared with tradition method, CCPSO is faster and has the higher precision. Therefore a new method is proposed.

Keywords: Chaos, Cultural algorithm, Particle Swarm Optimization, Logistic map.

1 Introduction

Particle swarm optimization (PSO) is a new global optimal evolution algorithm, which was proposed by Kennedy and Eberhart in 1995. It is inspired by the observations of the social behavior of animals, such as bird flocking, fish schooling and swarm theory [1]. And PSO has been successfully applied in different fields. However, it has some drawbacks, one of which is premature. In 1994, Reynolds proposed a cultural algorithm (CA) model which is focused on dividing the evolution space into population space and belief space, which enhances mutually during the evolution progress [2]. So it can improve the efficiency of the algorithm. In recent years, CA has drawn wide attention by many scholars, and has been got applied in many fields [3].

In this paper, we focus on the defect of PSO, and propose a new algorithm, namely, chaotic culture particle swarm optimization (CCPSO). In CCPSO, PSO is integrated into the framework of cultural algorithm model and utilized as the evolution method of population space. Meanwhile, chaos search operator is imported as the

^{*} Corresponding author.

knowledge structure of belief space. With the randomness and regularity of the chaos mechanism, CCPSO can prevent the search process from being trapped into the local optimal area. So the proposed algorithm can be more effective to get the global optimal solution. The performance of CCPSO has been demonstrated by several test functions. The test result shows the efficiency of the proposed algorithm. Finally, CCPSO is applied to the optimal operation of a cascade hydropower station. And satisfied result is obtained.

This paper is organized as followed. In section 2, we firstly introduced CA and PSO, then, on based of CA and PSO, CCPSO is proposed. In section 3, we test PSO and CCPSO with numerical experiments, and identified the effectiveness of the proposed algorithm. In section 4, we apply CCPSO to the problem of optimal operation of cascade hydropower stations. Finally, we conclude this paper in Section 5.

2 Chaos Cultural PSO

2.1 Cultural Algorithm

Derived from observing the cultural evolution process in nature, Robert G.Reynolds proposed a computational framework called cultural algorithm (CA). CA is mainly focused on genetic concepts and natural selection mechanism [2]. CA has been shown to be very effective in decreasing computational cost when combined with other algorithms [3].

The framework of CA is shown in Fig. 1, which consists of three components, a population space, a belief space and a protocol with which the two spaces exchange information. Any population based algorithm can be adopted as the evolution progress of the population space which produces knowledge [3]. Then, belief space selectively accepts the knowledge, and uses the knowledge to adjust its knowledge structure, with which the belief space can guide population space evolution in the next generation. The exchanging information of two spaces is implemented by the operation *accept()* and *influence()* according to the Communication Protocol [3].

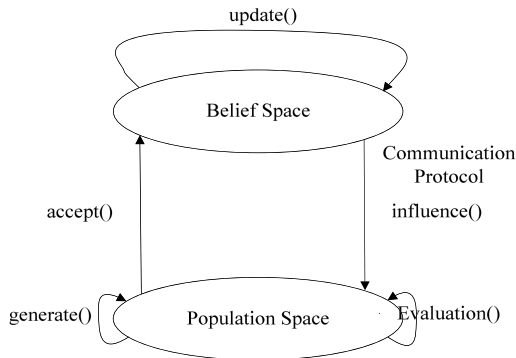


Fig. 1. Framework of CA

2.2 PSO

Particle swarm optimization (PSO) is a population-based Stochastic Optimization Algorithm. The rule of velocity and position updating is below:

$$v_{i,j}^{g+1} = w * v_{i,j}^g + c_1 * rand() * (Pbest_{i,j}^g - x_{i,j}^g) + c_2 * rand() * (Gbest_j^g - x_{i,j}^g) \quad (1)$$

$$x_{i,j}^{g+1} = x_{i,j}^g + v_{i,j}^{g+1}, i = 1, 2, \dots, Psize \quad j = 1, 2, \dots, D \quad (2)$$

Where, w is inertial factor; c_1, c_2 are acceleration factors; $x_{i,j}^g$ is the value of the j -th dimension of particle i in iteration g ; $rand()$ is a function which generates stochastic number in the range of $[0,1]$; $Pbest_{i,j}^g$ is the value of the j -th dimension of precious best position in iteration g . $Gbest_j^g$ is the value of the j -th dimension of global best position in iteration g . D is the dimension number.

The particles update their velocity and position repeatedly by equation (1) and (2) until the user-defined criterion condition is reached. The criterion condition can be the obtained while the maximal interactive number or the precision of solution is achieved [4].

2.3 CCPSO

PSO proposed a framework for solving complex optimal problems. However, it exhibits some drawbacks. One of which is premature. In this paper, chaos cultural particle swarm optimization (CCPSO) is proposed to solve the problem. The proposed algorithm inserts PSO into the framework of CA as the population space. Population space generates knowledge, then, Belief space accepts knowledge from population space by the operation *accept()* and guides evolution progress by the operation *influence()*. The operation *accept()* and *influence()* are designed according to the communication proctor to avoid the phenomenon of premature by dual-evolution and dual-promoting [3].

2.3.1 Knowledge Structure of Belief Space and Its Implementation

Chaos search operator is adopted to implement the knowledge structure of belief space. Chaos is a nonlinear phenomenon that widely exists in nature, which has good ability of local search. The Logistic map equation is:

$$\beta^{k+1} = \mu\beta^k(1 - \beta^k), k = 1, 2, \dots, \beta \in (0,1), \beta \neq 0.25, 0.5, 0.75 \quad (3)$$

Where, k is the iterative number. β^k is a stochastic number between 0 and 1. $\mu=4$.

Knowledge structure of belief space can be implemented by many ways [2]. In this paper, the belief space consists of elite individuals X^* and the chaotic search operator *ChaosSearch()*. *ChaosSearch()* disturbs the individuals in X^* to avoid the phenomenon of premature convergence.

The procedure of *ChaosSearch()* operate each individual in X^* is listed below:

Step1: Initialization. Initial D dimensional chaos variables $\beta^k = [\beta_1^k, \beta_2^k, \dots, \beta_j^k, \dots, \beta_D^k]^T$. $k=0$. Where, β_j^k is the j -th dimension of chaos variables.

Step2: Calculate $\beta^{k+1} = [\beta_1^{k+1}, \beta_2^{k+1}, \dots, \beta_j^{k+1}, \dots, \beta_D^{k+1}]^T$ according to equation (3).

Step3: Calculate Pc_j according to equation (4);

$$Pc_j = x_{j,\min} + \beta_j^{k+1}(x_{j,\max} - x_{j,\min}) \quad (4)$$

Where, Pc_j is variation scales. $x_{j,\max}$ and $x_{j,\min}$ are upper and lower limit.

Step4: Calculate the new position of current individual according to equation (5).

$$x_j^{k+1} = (1 - \lambda_g) \times x_j^k + \lambda_g \times Pc_j \quad (5)$$

Where, λ_g is shrinkage factor. x_j^k is the value of current particles at the j -th dimension.

Step5: Calculate new fitness of current position. $k=k+1$. Compare the new fitness with the previous one, if the new position's is better, replace the previous position with the new position.

Step6: If $k > \maxGen$ (The maximum iterative number, in this paper, it equals 150), end current searching, or turn Step2.

From equation (5), it can be concluded that λ_g determines the disturbance degree of chaos search operator to the elite individuals. In order to increase the efficiency of searching, λ_g is calculated according to equation (6).

$$\lambda_g = 1 - ((k - 1) / k)^m \quad (6)$$

Where, m is the parameter used to control the shrink velocity. From equation (6), we can draw a conclusion that at the beginning of searching, λ_g is large, chaos search operator searches globally to keep the diversity of population, while as g increases, λ_g become smaller, so that the precision of solution can be promoted by chaos searching in a small range around elite individuals.

2.3.2 The Implement of Communication Protocol

Accept() and *influence()* operation implement the communication of two populations, which improve the efficiency of optimal capacity of the algorithm. CCPSO executes the two operations every iterative 20 times in this paper. The procedure is as follows:

Accept(): Population space supplies belief space with the best N (N is the size of belief space) particles. Belief space compares the fitness of the particles with its own ones. Then, the N optimal particles are selected as the new population of belief space.

Influence(): Belief space supplies population space with its optimal $0.5N$ particles. Population space compares the $0.5N$ particles with its own particle's fitness. Then, the population space abandons the worst $0.5N$ particles.

2.3.3 Procedure of CCPSO

Step1: Initialization. Initialize population space, belief space and parameters. Set $g=0$.

Step2: Evolve population space. Calculate and update each individual's velocity and position according to equation (1) and (2).

Step3: Belief space utilizes the knowledge structure which is described in chapter 2.3.1. And evolves with the chaotic search operator.

Step4: If *FlagA* is true, *accept()* operation is implemented, otherwise turn Step5;

Step5: If *FlagI* is true, *influence()* operation is implemented, otherwise turn Step6;

Step6: If $g \geq \text{MaxGen}$ (the maximum number of evolution), export the optimal solution as the final solution. Or, $g=g+1$, then, turn Step2;

Where, the size of belief space is set to 10% of the population space. *FlagA* and *FlagI* are the identifications whether *accept()* or *influence()* operation is implemented.

3 Benchmark Function Tests

In order to test the effectiveness of CCPSO, we use four well-known functions: The Rosenbrock function, the Griewank function, the Rastrigin function and the Sphere function. The global optimum solution of all functions is known to be 0.

(1) Rosenbrock

$$f(x) = \sum_{i=1}^{n-1} (100(x_{i+1} - x_i^2) + (x_i - 1)^2), x_i \in [-30, 30] \quad (7)$$

(2) Griewank

$$f(x) = 1 + \sum_{i=1}^N \frac{x_i^2}{4000} - \prod_{i=1}^N \cos\left(\frac{x_i}{\sqrt{i}}\right), x_i \in [-600, 600] \quad (8)$$

(3) Rastrigin

$$f(x) = \sum_{i=1}^N [x_i^2 - 10 \cos(2\pi x_i) + 10], x_i \in [-5.12, 5.12] \quad (9)$$

(4) Sphere

$$f(x) = \sum_{i=1}^N x_i^2, x_i \in [-5.12, 5.12] \quad (10)$$

For all functions, the parameter of PSO is $w=0.729$, $c_1=c_2=2$. The functions are tested with 30 variables, and iterative number is set separately with 500 and 1000. In order to eliminate the influence of randomness, each function is independently processed 50 times. The test results are shown in Table 1.

Table 1. The result of CCPSO

Generation	function	Algorithm	optimal	average	average time(ms)	
Gen=500	Rosenbrock	PSO	11.98	21.3	1150	
		CCPSO	1.42E-06	1.32E-04	1243	
	Griewank	PSO	9.86E-03	2.82E-02	940.7	
		CCPSO	6.69E-13	8.73E-12	1240.6	
	Rastrigin	PSO	29.85	40.69	1067.2	
		CCPSO	1.78E-15	7.55E-13	1046.9	
	Sphere	PSO	3.24E-10	1.20E-09	715.8	
		CCPSO	1.12E-15	8.88E-15	765.6	
	Gen=1000	Rosenbrock	PSO	11.72	20.03	2134.4
			CCPSO	1.50E-07	2.80E-06	2334.6
Griewank		PSO	7.40E-03	2.80E-02	1824.9	
		CCPSO	1.11E-16	1.74E-16	1961.8	
Rastrigin		PSO	28.85	39.6	2014	
		CCPSO	1.78E-15	2.00E-15	2037	
Sphere		PSO	4.67E-24	1.42E-22	1465.5	
		CCPSO	1.03E-28	2.47E-30	1515.8	

From Table 1, it can be concluded that CCPSO has better accuracy than PSO in these four test functions with the same calculate time. In the test of first three functions, as the iterative number rises from 500 to 1000, the accuracy of PSO does not increased much, which means PSO traps into local optimal, the accuracy cannot get a further improvement. However, the accuracy of CCPSO is increased expressly. Consequently, CCPSO can avoid premature effectively and get a better convergence precision.

4 Optimal Operation of Cascade Hydropower Stations Based on CCPSO

The optimal operation of cascade hydropower station is a typical problem in the optimal fields of hydropower energy system. The operation purpose is to schedule the power discharge of each scheduling period, under the condition of satisfying constraint which can get the maximum power generation benefit during the scheduling period. Because of the complex power and hydraulic relation between cascade power systems, the optimal operation of cascade hydropower station is a large scale, dynamic, and strong coupling nonlinear model, which is difficult to get the global optimal solution. In recent years, lots of relative algorithms have been researched. Such

as, Dynamic Programming (DP), Progressive Optimality Algorithm (POA), Genetic Algorithm (GA), Particle Swarm Optimization (PSO) and so on. And many achievements have also been obtained. However, all of the methods listed above exists some defects. Though DP can solve the optimal operation of single reservoir, while being used in the optimal operation of cascade hydropower station, it cause the problem of “curse of dimensionality”; POA is sensible to the initial Solution; GA and PSO can get a reasonable solution, but it is easily trapped in the local optimal, and difficult to solve complex constraints in practice.

4.1 Objective Function

The object is to maximize the total generated energy. In this paper, Period of Time is month, and fitness function is generated energy. As is described below:

$$E = \max \sum_{i=1}^N \sum_{t=1}^T A_i \cdot H_{i,t} \cdot Q_{i,t} \cdot M_t \quad (11)$$

Where, E is the max generated energy; N is the number of power station. A_i is output power coefficient. $Q_{i,t}$ is the discharge of average power. $H_{i,t}$ is the head of i -th power station. T is the total number of dispatching period's time. And M_t is the length of period's time.

4.2 Constraints

(1) Water level upper and lower limit

$$Z_{i,t}^{\min} \leq Z_{i,t} \leq Z_{i,t}^{\max} \quad (12)$$

(2) Power output upper and lower limit

$$N_{i,t}^{\min} \leq N_{i,t} \leq N_{i,t}^{\max} \quad (13)$$

(3) Discharge upper and lower limit

$$Q_{i,t}^{\min} \leq Q_{i,t} \leq Q_{i,t}^{\max} \quad (14)$$

(4) Water reservoir balance

$$V_{i,t+1} = V_{i,t} + (I_{i,t} - Q_{i,t}) \cdot M_t \quad (15)$$

(5) Hydraulic connection of cascade reservoirs

$$I_{i,t} = Q_{i-1,t-\tau_{i-1}} + q_{i,t} \quad (16)$$

Where, $Z_{i,t}^{\max}$ and $Z_{i,t}^{\min}$ are the upper and lower limit of water level. $N_{i,t}^{\max}$ and $N_{i,t}^{\min}$ are the upper and lower limit of power output. $Q_{i,t}^{\max}$ and $Q_{i,t}^{\min}$ are the upper and lower limit of discharge. $I_{i,t}$ is the i -th power station's inflow runoff at the t -th period. $Q_{i-1,t-\tau_{i-1}}$ is the outflow of upstream station. τ_{i-1} is the time that water flows from the

$i-1$ -th power station to i -th power station. $q_{i,t}$ is the local inflow of the i -th power station at t period.

4.3 Encoding and the Initialization of Original Population

In this paper, water level is adapted as the optimal variable. When applying CCPSO to solve cascade hydropower stations optimal operation problem, optimal vector of each individual is described by water level series $H = \{h_{1,1}, h_{1,2}, \dots, h_{1,T}, \dots, h_{N_h,1}, h_{N_h,2}, \dots, h_{N_h,T}\}$, which represents the water level of each power station at each period. Then generate an initial feasible population in the feasible region. Where, the j -th individual is initialed as $H^j = \{h_{1,1}^j, h_{1,2}^j, \dots, h_{1,T}^j, \dots, h_{N_h,1}^j, h_{N_h,2}^j, \dots, h_{N_h,T}^j\}$.

4.4 Constraints Handling

Constraints of cascade hydropower stations optimal operation are complex. The most popular method to deal with constraints is penalty function, however, penalty function has to try many times to calculate the penalty factor, and this increases the computational complexity of optimization. Aiming at this problem, the method of dealing with constraints in [10] is imported in this paper. The method focus on transforming constraints to water level constrains in every period according to the feature of cascade hydropower stations optimal, then guides evolution of CCPSO by the range of the water level constrains. So the original problem is transformed into a non-constrain optimal problem. More details can be found in [10].

4.5 Fitness Function

CCPSO searches in the feasible region, while after each evolution; the individuals are also feasible, so the penalty is not necessary. Equation (17) is used as the fitness function in this paper.

$$f(H) = E = \max \sum_{i=1}^N \sum_{t=1}^T A_i \cdot H_{i,t} \cdot Q_{i,t} \cdot M_t \quad (17)$$

4.6 Result and Analysis

In order to identify the feasibility and effectiveness of CCPSO, we calculate with a cascade hydropower station which includes two stations (The upstream station A and the downstream station B). The process of optimization is used separately with CCPSO and POA, where POA is a tradition method which is widely used in the optimal operation of cascade hydropower station.

The parameters of CCPSO and constraints are shown in Table 2 and Table 3. The result of operation is shown in Table 4. The result of comparing with CCPSO and POA is shown in Table 5.

Table 2. Parameters of CCPSO

name	w	c_1	c_2	$Psize$
value	0.729	2.0	2.0	100

Table 3. Constraints of the cascade hydropower station

Parameter name	A	B
Discharge volume	[1580,98800]	[3200,86000]
Upstream water level	[145,175]	[63.0,66.5]
Downstream water level	[63.0,71.8]	[38.0,58.6]
Output	[499,1820]	[104,271.5]
Output power coefficient	8.5	8.4
Priming level of regulation	175	65

The Table 4 illustrates that: compared with the result of POA, CCPSO processes faster and has better optimal operation solution. The total generated energy calculated by CCPSO is $0.3242(10^8 \text{kwh})$ more, besides, the time cost is halved. Since POA is sensitive with the original solution, the range of searching is limited, and when in actual application, a reasonable original solution is always difficult to obtain, however, CCPSO is not sensitive with the original solution, with the comprehensiveness of chaos search, the precision of solution can get a higher increase.

Table 4. The result of CCPSO

Power station	month	inflow	Z	Q	N	Sp
		(m^3/s)	(m)	(m^3/s)	(10^4kW)	($10^8 \text{m}^3/\text{s}$)
	1	4460	175	5378	499.001	0
	2	4218	172.5369	5528	499.001	0
	3	5525	169.09	5628	498.999	0
	4	7516	168.7874	5492	499	0
	5	8301	174.2329	16282	1300.096	0
A	6	22113	145	22113	1473.774	0
	7	35387	145	35387	1675.508	9482
	8	25867	145	25867	1710.657	0
	9	26136	145	26136	1727.405	0
	10	21532	145	13262	1067.578	0
	11	10634	175	10634	994.218	0
	12	6622	175	6622	620.833	0

Table 4. (Continued)

B	1	4460	65	5378	107.591	0
	2	4218	64.1	5528	105.946	0
	3	5525	63.1	5628	102.817	0
	4	7516	63.09	5492	100.704	0
	5	8301	63.15	16282	231.103	0
	6	22113	63.03	22113	261.953	0
	7	35387	63.1	35387	271.5	6743
	8	25867	63.08	25867	271.5	429
	9	26136	63.9	26136	271.5	637
	10	21532	63.6	13262	205.713	0
	11	10634	64.1	10643	179.412	0
	12	6622	64.6	6622	125.651	0
E(10^8 kwh)	A	921.2539	B	163.852	processing time(ms)	8651
	total		1085.106			
Sp(10^8 m ³)	A	253.9636	B	205.6099		
	total		459.5735			

Table 5. The comparing of CCPSO and POA

Algorithm	variables	A	B	TOTAL	Processing time(ms)
CCPSO	E(10^8 kwh)	921.2539	163.852	1085.106	8651
	Sp(10^8 m ³)	253.9636	205.6099	459.5735	
POA	E(10^8 kwh)	921.2049	163.5769	1084.7818	16336
	Sp(10^8 m ³)	253.9636	205.6092	459.5728	

5 Conclusions

In this paper, focused on the premature of PSO, we proposed an algorithm called Chaos Cultural PSO. CCPSO adopts PSO in the framework of CA, and combines with chaos search as the knowledge structure of belief space in the framework. By the communication of two populations, CCPSO enhance the searching efficiency of algorithm. The result of test functions shows that CCPSO can avoid premature. And has the better performance than PSO. Finally, we applied CCPSO to cascade hydropower stations optimal operation. Compared CCPSO with tradition algorithm, the results verify its superiority in both efficiency and precision. Therefore a new way is proposed to resolve cascade hydropower stations optimal operation.

Acknowledgements

This paper is supported by the projects of Natural Science Foundation of China (No. 50539140), the Special Research Foundation for the Public Welfare Industry of

the Ministry of Science and Technology and the Ministry of Water Resources (No. 200701008) and the projects of Natural Science Foundation of Hubei province (No. 2008CDA088).

References

1. Kennedy, J., Eberhart, R.C.: Particle Swarm Optimization. In: Proc. IEEE Int. Conf. Neural Networks, vol. IV, pp. 1942–1948. IEEE Service Center, Los Alamitos (1995)
2. Reynolds, R.G.: An introduction to cultural algorithms. In: Proceedings of the 3rd annual Conference on Evolution Programming, San Diego, California, pp. 131–136 (1994)
3. Reynolds, R.G., Mostafa, Z., Ali, T.J.: Mining the Social Fabric of Archaic Urban Centers with Cultural Algorithms. *IEEE Computer* 41(1), 64–72 (2008)
4. Yang, X.M., Yuan, J.S., Yuan, J.Y., Mao, H.N.: A modified particle swarm optimizer with dynamic adaptation. *Applied Mathematics and Computation* 189, 1205–1213 (2007)
5. Carlos, A.C.C., Lechuga, M.S.: MOPSO: A Proposal for Multiple Objective Particle Swarm Optimization. In: IEEE Congress on Evolutionary Computation (CEC 2002), Honolulu, Hawaii, USA, pp. 1051–1056 (2002)
6. Paterlini, S., Krink, T.: Differential evolution and particle swarm optimization in partial clustering. *Computational Statistics & Data Analysis* 50(5), 1220–1247 (2006)
7. Liu, B., Wang, L., Jin, Y.H., Tang, F., Huang, D.X.: Improved particle swarm optimization combined with chaos. *Chaos Solutions & Fractals* 25, 1261–1271 (2005)
8. Yuan, X.H., Cao, B., Yuan, Y.B.: Hydrothermal scheduling using chaotic hybrid differential evolution. *Energy Conversion and Management* 49, 3627–3633 (2008)
9. Yuan, X., Yuan, Y.: A hybrid chaotic genetic algorithm for short-term hydro system scheduling. *Mathematics and Computers in Simulation* 59, 319–327 (2002)
10. Yang, J.J.: Joint Optimal Regulation for Cascade Hydropower Stations based on MOPSO and Set Pair Analysis Decision-making Approach (in Chinese). Ph.D thesis, Huazhong University of Science & Technology, Wuhan (2007)

Application of Visualization Method to Concrete Mix Optimization

Bin Shi¹, Liexiang Yan², and Quan Guo²

¹ School of Mechatronic Engineering, Wuhan University of Technology,
Wuhan, 430070, China

² School of Chemical Engineering, Wuhan University of Technology,
Wuhan, 430070, China
sbin125@gmail.com

Abstract. Due to the complex interaction of components, the design of concrete mix becomes difficult. This paper presents an artificial neural network (ANN) based visualization method to optimize the concrete mix design. It aims to minimize the cost of concrete such that all desired qualities are maintained. The procedure can be described as mapping data of concrete mix from multidimensional space to a two-dimensional plane with an ANN model, and then generating concrete property contours on this plane. The optimized mix proportions region can be determined intuitively based on the contours distribution. By means of an inversion mapping algorithm, the optimal point in this region can be mapped inversely to the original multidimensional space. Practical production test results show that good concrete mixes, which agree with the concrete compressive strength criterion and have lower cost, can be obtained. Application of this method can contribute significant benefits to the commercial concrete companies.

Keywords: ANN, Visualization Method, Concrete Mix, Optimization.

1 Introduction

Concrete is the most widely used structural material for construction today. Based on the strong increasing demand for improving concrete performance, many admixtures and additives are added into concrete, which makes the interactions between various components become more complex than before. Since the highly nonlinear relationships exist between components and concrete properties, it is difficult to set up any mathematical model to take all factors into account, therefore, the traditional methods that build calculation models for concrete mix design have been difficult to meet the design requirements. How to optimize concrete mix design has become a focus in recent years.

Many works have been done on the optimization for concrete mix design, which Bin Chen has adopted stepwise regression method to optimize concrete mix [1]; I-Cheng Yeh, Kim Jong-In, Mohammed H Alawi and Ji Tao have proposed ANN methods in concrete mix design [2-5]; Bai Y and Zain MFM have established an expert system for concrete mix design respectively, and the systems' selection of

concrete mix were both compared favourably with those of experts [6-7]. Though all of the above methods can achieve a satisfactory fitting between predicted results and experimental results, they didn't consider controlling the raw materials cost simultaneously, which is in fact very important for practical use. This paper applies the ANN based visualization method (VM) to optimize concrete mix design, and determine an optimal mix proportion, which meets to the certain performance requirements and has lower cost. The effectiveness of this method has been tested on practical production.

2 The Visualization Method

2.1 Basic Principles

The VM is an effective method for finding an optimized operating region and an optimal operating point based on processing practical production data or experimental data. The basic principles behind this method are shown in Fig.1. Firstly, the sample data in multidimensional space are mapped to a two-dimensional plane with a mapping model; meanwhile, the contours of the objective function or functions are generated automatically in this plane. Then, the optimized operating direction or region can be located intuitively according to the contours distribution. Finally, a point found in this region, which, although not strictly optimal, is near-optimal, can be mapped back to the original multidimensional space with an inversion mapping method, and will be represented in terms of original variables. It is beneficial as a guide for practical production and scientific experiments [8].

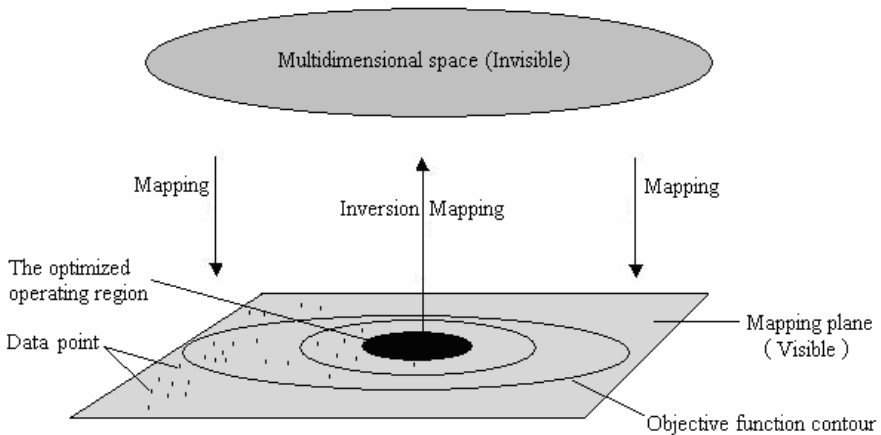


Fig.1. Principles of visualization method

2.2 Mapping Model

The mapping relationship between multidimensional space and mapping plane is established based on an ANN shown in Figure.2. The information transfer for this network is shown as follows:

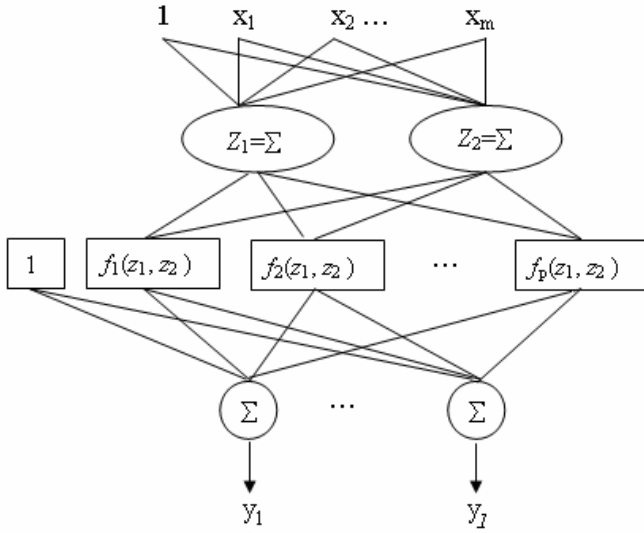


Fig. 2. Mapping model

From inputs to mapping plane:

$$z_1 = w_1 X^T, z_2 = w_2 X^T . \tag{1}$$

From mapping plan to outputs:

$$Y = VP^T = \begin{bmatrix} v_{10} & v_{11} & \dots & v_{1p} \\ v_{20} & v_{21} & \dots & v_{2p} \\ \dots & \dots & \dots & \dots \\ v_{l0} & v_{l1} & \dots & v_{lp} \end{bmatrix} \begin{bmatrix} 1 \\ f_1(z_1, z_2) \\ \dots \\ f_p(z_1, z_2) \end{bmatrix} . \tag{2}$$

where:

$$\begin{aligned} X &= [1, x_1, x_2, \dots, x_m] & w_j &= [w_{0j}, w_{1j}, \dots, w_{mj}] & j &= 1, 2 \\ Y &= [y_1, y_2, \dots, y_l] & P &= [1, f_1(z_1, z_2), \dots, f_p(z_1, z_2)] \\ V &= [v_1, v_2, \dots, v_l]^T_{l \times p} & v_k &= [v_{k0}, v_{k1}, v_{k2}, \dots, v_{k(p+1)}] & k &= 1, 2, \dots, l \end{aligned}$$

where X is an m -dimensional input vector, Y is an l -dimensional output vector, z_1 and z_2 are two variables on mapping plane, w_1, w_2 and V are the weight vectors of the neural network, P is a nonlinear extending vector for enhancing mapping effect, and in this work we take simply $P = [1 f_1(z_1, z_2) f_2(z_1, z_2) \dots f_p(z_1, z_2)] = [1 z_1 z_2 z_1^2 z_2^2 z_1 z_2]$.

The determination of the weight vectors for the network can be translated into solving the following unconstraint non-convex nonlinear programming problem:

$$E = \min \sum_{t=1}^n \sum_{k=1}^l |d_k(t) - y_k(t)| \quad (3)$$

where n is the number of samples, $d_k(t)$ and $y_k(t)$ are practical value and network output value of the k -th function corresponding to the t -th sample, respectively. In order to obtain the weight values of the network, line-up competition algorithm (LCA) is adopted to solve above nonlinear programming problem.

2.3 Inversion Mapping Algorithm

After the sample data in multidimensional space have been mapped to a plane and the contours of objective functions have been produced, it is easy to locate the optimized region or optimal point on the mapping plane according to the contours distribution of objective functions. How to inverse the optimal point to the multidimensional space and to be represented with the original variables is an important problem for practical use. The following theorem can solve the inversion mapping problem.

$$x^c = x^a + \beta(x^b - x^a) \quad (4)$$

Where, x^a and x^b are the points in multidimensional space relative to the points a and b on the plane; x^c is the point in multidimensional space relative to the point c on the plane, which c is any point of the line through points a and b; β is step size and its value equals the ratio of the distance between points a and c to the distance between points a and b.

$$\beta = \frac{\overline{ca}}{\overline{ba}} = \frac{z_1^c - z_1^a}{z_1^b - z_1^a} \quad (5)$$

where $0 \leq \beta \leq 1$ represents interpolation, $\beta \geq 1$ for extrapolation.

Using this formula, given an optimal point on mapping plane, the corresponding optimized point in original multidimensional space can be obtained easily.

2.4 The Features of Visualization Method

The dimension reduction from multidimensional space to a two-dimensional plane was achieved by the combination of ANN and LCA. The ANN was adopted to be mapping model and the LCA was used to determine the model parameters. Based on the excellent global optimization performance of LCA, the nonlinear mapping and the fitting capacity of mapping model can be enhanced.

Through introducing the inversion mapping algorithm, the problem that points in two-dimensional plane can't be returned to the original multidimensional space after dimension reduction, which exists in multivariate statistical methods such as principal component analysis (PCA), and pattern recognition, was solved successfully. It made the VM more valuable in practical use.

The contours of several objective functions can be generated simultaneously on a plane with the VM. Under this situation, a satisfactory solution which gives a balance between the alternative objective functions can be obtained intuitively. It has the equally effectiveness for optimization problems with constraints.

3 Practice Applications

In concrete mix design, the compressive strength of concrete is regarded as the most important property. Many other properties, such as elastic modulus, water tightness or impermeability, resistance to weathering agents, etc. are directly related to the compressive strength. So, this paper takes the 28th day compressive strength, which is universally accepted as a general index of concrete strength, as the main mix design target. The aim of work is to optimize the C30 and C25 grade concrete cost under a certain compressive strength criterion for a commercial concrete company.

A group of experimental data was collected from the company. The experimental scheme and results are list in Table 1. There are eight important factors influencing the concrete cost(y_1) and compressive strength(y_2), namely, cement(x_1), fly ash(x_2), mineral powder(x_3), sand(x_4), water reducer(x_5), water(x_6), blue-stone(x_7) and red-stone (x_8).

Table 1. Uniform experimental scheme and results

NO.	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	y_1 (yuan/M ³)	y_2 (MPa)
1	179.8	63.6	78	755.4	8.47	171.2	518.6	639	155.38	32
2	196	73.5	103.6	727.6	8.6	177.7	437.4	707.6	165.35	34.2
3	199.7	78.8	75.1	795.3	8.18	173.7	334.4	723.8	158.03	31.8
4	201.1	82.9	93.5	712.9	7.71	169.8	284.5	853.6	161.54	36.8
5	220.9	94.6	72.4	694.3	7.92	175	536.6	605.1	166.73	33.8
6	220.5	97.7	90.6	750.3	8.81	178.1	420.2	619.9	170.05	34.9
7	231.1	59.1	66.3	702.4	7.29	177.7	395.1	773.9	163.84	34.7
8	228.6	63.1	83.5	750.6	6.77	172.6	288.8	773	163.17	38.2
9	254.2	75.0	64	749.9	7.08	176	533.2	550.5	171.67	36.2
10	251.3	78.5	82	663.2	6.59	176.6	490.7	661.2	173.26	40.7
11	253.9	83.4	56.2	729.3	6.26	177.4	386.4	687	166.54	39.8
12	270.4	93	79	699	6.28	184.6	309.8	743.9	175.49	37.9

3.1 Mapping Results

The data processing results obtained with the VM are listed in Table 2. It is clear that the maximal absolute value of relative error between real results and calculated results is no more than 0.52. It shows that the fitting between calculated values and real values is satisfactory, and the mapping relations can be used to predict the concrete mix.

Table 2. Export value of the mapping model and the actual value comparison

NO.	Real concrete cost	Calculated concrete cost	Relative error	Real concrete strength	Calculated concrete strength	Relative error	Mapping plane coordinate	
							z1	Z2
1	155.38	155.08	0.1934	32	31.885	0.3607	-0.55393	-0.45415
2	165.35	165.86	-0.3075	34.2	34.191	0.0263	0.25474	-0.81582
3	158.03	158.4	-0.2336	31.8	31.953	-0.4788	-0.3983	-0.35515
4	161.54	161.6	-0.0371	36.8	36.774	0.0707	0.15647	-1.1181
5	166.73	167.14	-0.2453	33.8	33.874	-0.2185	0.08599	-0.037199
6	170.05	170.22	-0.0999	34.9	34.957	-0.1631	0.42074	-0.67403
7	163.84	163.55	0.1773	34.7	34.664	0.1039	-0.13353	-0.024312
8	163.17	163.15	0.0123	38.2	38.232	-0.0837	0.27394	-1.0998
9	171.67	172.56	-0.5158	36.2	36.093	0.2965	0.63094	-0.43969
10	173.26	173.08	0.1040	40.7	40.69	0.0246	0.83257	-0.59662
11	166.54	166.39	0.0901	39.8	39.723	0.1938	0.07596	0.33055
12	175.49	175.36	0.1711	37.9	37.967	-0.1765	0.7413	0.16516

Figure.3 shows the mapping diagram for this problem. Sample data of concrete mix in multidimensional space are mapped and reduced dimension to a plane, 12 black points in figure represent 12 groups of concrete mix experiments respectively. The contours of concrete compressive strength and cost are generated automatically on this plane. The real lines represent the contours of concrete cost of every cubic

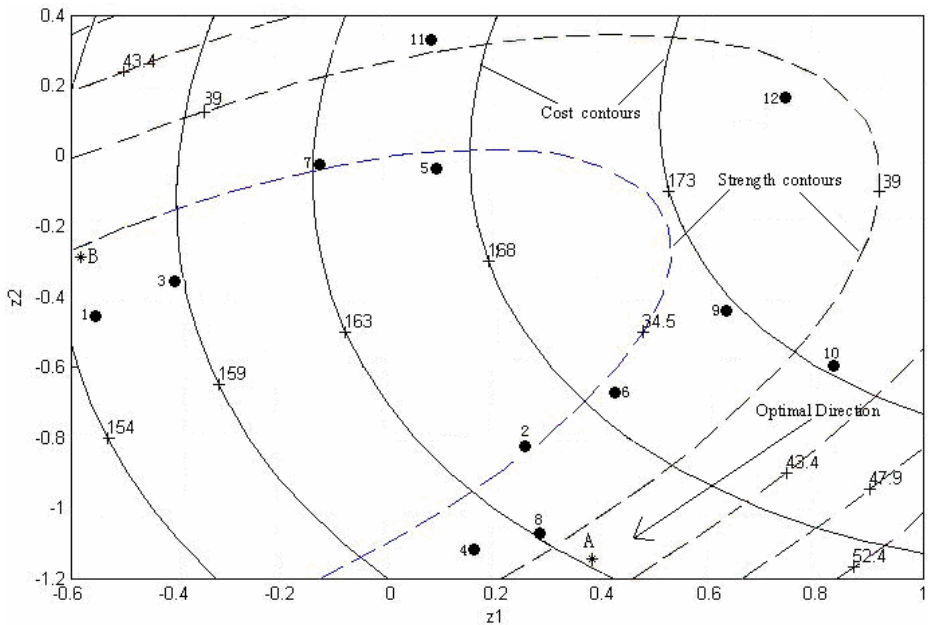


Fig.3. Mapping figure

meter and the dash-dotted lines are the contours of concrete strength of 28th day. The optimal mix proportions can be determined intuitively based on contours distribution. By means of the inversion mapping algorithm, optimal point in this plane can be mapped inversely to original multidimensional space and represented in terms of practical mix proportion data.

3.2 Prediction and Verification for the Optimal Mix Proportion

Based on the criterion for C30 concrete, the 28th day concrete compressive strength must be controlled between 38MPa to 42MPa. And the current concrete cost of C30 is 166 Yuan/m³. It can be seen from Figure.3 that the concrete cost will decrease, and the concrete strength will be controlled between 38MPa to 42MPa in the direction of the arrow. Taking points 1 and 8 as references and step size as 1.1, a predicted point asterisk A is obtained through extrapolation in the direction of the arrow, the concrete cost and strength at the predicted point are 163.2 Yuan/m³ and 41MPa respectively. The corresponding mix proportion parameters are list in Table 3. For C25 concrete, the 28th day concrete compressive strength must be controlled between 30MPa to 34MPa, and the current concrete cost is 156 Yuan/m³. The asterisk B in Figure.3 is a predicted point obtained by extrapolating from points 6 to 3 ($\beta = 1.2$). The corresponding mix proportion parameters are list in Table 3, too.

Table 3. Prediction results

Samples	β	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	y_1	y_2
A: 1→8→	1.1	233.5	63.06	84.08	750.1	6.61	172.7	265.9	786.4	163.2	41
B: 6→3→	1.2	195.6	74.97	72	804.3	8.058	172.8	317.2	744.6	155.2	33.8

→: Extrapolating direction, β : step size.

Table 4. Comparison of test results and prediction results

samples	Compressive strength(MPa)			Cost(Yuan/m ³)		
	Predicted	Test	Errors	Predicted	Test	Errors
C30	41	42.5	3.53	163.2	163.7	0.31
C25	33.8	34	0.59	155.2	155.8	0.39

The predicted concrete mixes have been tested on a group of verification experiments. Table 4 shows the comparison of the test results and prediction results. In the compressive strength, the values of error percent differences between the test and the predicted results compared to the test values are 0.59% and 3.53% for C30 and C25 grade concrete respectively. And in the concrete cost, the values of error percent differences between the test and the predicted results compared to the test values are 0.31% and 0.39%. Therefore, the compressive strengths of concrete predicted by the proposed method agree with those resulting from compressive tests, at the same time, the concrete costs reduce markedly, which demonstrate the effectiveness of the VM.

4 Conclusions

This paper presented the application of VM to optimize concrete mix. The concrete mix data from a commercial concrete company were processed by the VM. The compressive strength and the concrete cost contours were mapped on a plane simultaneously, and the optimal mix proportions region can be determined intuitively according to the contours distribution. Through extrapolation or interpolation with different step sizes in the optimal region, an optimal mix proportion point can be located. In this study, the maximal error between real results and calculated results was found to be 0.52 during the data processing by the VM. The validity of the proposed method was proven by comparing the predicted compressive strength and cost with the practice production test results of the company. The maximum errors between the predicted and tested results were 3.53% in the compressive strengths and 0.59% in the concrete cost.

This study demonstrated the effectiveness of the VM in optimization the compressive strength and the cost of concrete based on practical concrete mix parameters. Application this method can contribute significant benefits to the commercial concrete companies. As a future study, other important factors that also affect the concrete strength, such as uncertainty of raw materials will be considered in the VM. Then the VM will become more effective and the optimization results will become more accurate and reliable.

References

1. Chen, B., Li, F.Q., Liu, G.H., et al.: Study on Nonlinear Multi-objective Optimization Algorithm for Concrete Mix Proportions. *J. Zhejiang Univ. (Eng. Sci.)* 39, 16–19 (2005) (in Chinese)
2. Yeh, I.C.: Design of High Performance Concrete Mixture Using Neural Networks and Nonlinear Programming. *J. Comp. in Civ. Eng.* 13, 36–42 (1999)
3. Jong-In, K., Doo, K.K., Maria, Q.F., Frank, Y.: Application of Neural Networks for Estimation of Concrete Strength. *J. Mat. in Civ. Eng.* 16, 257–264 (2004)
4. Mohammed, H.A., Maher, I.R.: Applications of neural network for optimum asphaltic concrete mixtures. *WSEAS Trans. Inf. Sci. Appl.* 2, 1913–1917 (2005)
5. Ji, T., Lin, T.W., Lin, X.J.: A Concrete Mix Proportion Design Algorithm Based on Artificial Neural Networks. *Cem. Conc. Res.* 36, 1399–1408 (2006)
6. Bai, Y., Amirhanian, S.N.: Knowledge-based Expert System for Concrete Mix Design. *J. Cons. Eng. Mng.* 120, 357–373 (1994)
7. Zain, M.F.M., Islam, M.N., Basri, I.H.: An Expert System for Mix Design of High Performance Concrete. *Adv. in Eng. Software* 36, 325–337 (2005)
8. Yan, L.X., Bogle, I.D.L.: A Visualization Method for Operating Optimization. *Comp. Chem. Eng.* 31, 808–814 (2007)

A Novel Nonparametric Regression Ensemble for Rainfall Forecasting Using Particle Swarm Optimization Technique Coupled with Artificial Neural Network

Jiansheng Wu¹ and Enhong Chen²

¹ Department of Mathematics and Computer, Liuzhou Teacher College
Guangxi, Liuzhou, China
wjsh2002168@163.com

² Department of Computer, University of Science and Technology of China
Hefei, Anhui, China

Abstract. In this study, we propose a novel nonparametric regression (NR) ensemble rainfall forecasting model integrating generalized particle swarm optimization (PSO) with artificial neural network (ANN). First of all, the PSO algorithm is used to evolve neural network architecture and connection weights. The evolved neural network architecture and connection weights are input into a new neural network. The new neural network is trained using back-propagation (BP) algorithm, generating different individual neural network. Then, the principal component analysis (PCA) technology is adopted to extract ensemble members. Finally, the NR is used for nonlinear ensemble model. Empirical results obtained reveal that the prediction by using the NR ensemble model is generally better than those obtained using other models presented in this study in terms of the same evaluation measurements. For illustration and testing reveal that the NR ensemble model proposed can be used as an alternative forecasting tool for a Meteorological application in achieving greater forecasting accuracy and improving prediction quality further.

Keywords: Nonparametric Regression, Neural Network Ensemble, Particle Swarm Optimization, Rainfall Forecasting.

1 Introduction

Accurate forecasting of rainfall has been one of the most important issues in hydrological research, because early warnings of severe weather, made possible by timely and accurate forecasting can help prevent casualties and damages caused by natural disasters. In general, rainfall forecasting involves a rather complex nonlinear data pattern, for example pressure, temperature, wind speed and its direction, meteorological characteristics of the catchments and so on [1]. Although a physically-based approach for rainfall forecasting has several advantages, given the short time scale, the small catchments area, and the massive

costs associated with collecting the required meteorological data, it is not a feasible alternative in most cases because it involves many variables which are interconnected in a very complicated way, and the volume of rainfall calculation require sophisticated mathematical tool [2, 3]. Recurrent Artificial Neural Network (ANN) have played a crucial role in forecasting rainfall data. [4, 5]. ANN is based on a model of emulating the processing of human neurological system to find out related spatial and temporal characteristics from the historical rainfall patterns (especially for nonlinear and dynamic evolutions) [6, 7]. Due to without understanding the physical laws and any assumptions of traditional statistical approaches required, ANN is widely applied to solve hydrological problems including rainfall forecasting [8].

The application of an ANN, however, involves a complicated development process. As ANN approaches want of a rigorous theoretical support, effects of applications strongly depend upon operator's experience. In the practical application, the results of many experiments have shown that the generalization of single neural network is not unique. That is, ANN results are not stable. Even for some simple problems, different structures of neural networks (e.g., different number of hidden layers, different hidden nodes and different initial conditions) result in different patterns of network generalization. If carelessly used, it can easily learn irrelevant information (noises) in the system (over-fitting) and limit applications of ANN in the practical application [9, 10]. In order to overcome the main limitations of ANN, recently a novel ensemble forecasting model, i.e. neural network ensemble (NNE), has been developed. Because of combining multiple neural networks learned from the same training samples, NNE can remarkably enhance the forecasting ability and outperform any individual neural network. It is an effective approach to the development of a high performance forecasting system [11].

In general, NNE is constructed in two step, i.e. training a number of individual neural network and then combining the component predictions. Different from the previous work, this study proposes a novel NR ensemble rainfall forecasting method in terms of PSO technique coupled with ANN (NR-PSO-ANN). PSO algorithm is applied to evolve neural network architecture and connection weights. The evolved neural network architecture and connection weights are input into a new neural network. The new neural network is trained using back-propagation (BP) algorithm. The output is obtained by NR. The rest of this study is organized as follows. Section 2 describes the building process of the NR ensemble rainfall forecasting model in detail. For further illustration, this work employs set up a prediction model for daily mean field of circulation and daily rainfall in Guangxi are used for testing in Section 3. Finally, some concluding remarks are drawn in Section 4.

2 The Building Process of the NR Ensemble Model

In this section, a triple-phase nonlinear neural network ensemble model is proposed for rainfall forecasting. First of all, many individual neural predictors are

generated by PSO technique. Then an appropriate number of neural predictors are selected from the considerable number of candidate predictors. Finally, selected neural predictors are combined into an aggregated neural predictor by NR.

2.1 Generating Individual Neural Network Predictors

Recently, a new evolutionary computation technique, the PSO is applied. Its development was based on observations of the social behavior of animals such as bird flocks, fish choosing, and swarm theory. Each individual in PSO is assigned with a randomized velocity according to its own and its companions' flying experience. The individuals, called particles, are then flown through hyperspace [12, 13]. Position-speed relation model of PSO operates easily. The method of using PSO to evolve neural networks includes three steps: (i) using global searching ability of PSO to find an appropriate network architecture and connection weights; (ii) using BP algorithm to search peak value(s) in detail; (iii) obtaining individual neural network.

Mathematically, optimization problems of PSO-Neural Network can be described as follows:

$$\left\{ \begin{array}{l} \min E(w, v, \theta, r) = \frac{1}{N_1} \sum_{k=1}^{N_1} \sum_{t=1}^n [y_k(t) - \hat{y}_k(t)] < \varepsilon_1 \\ \hat{y}_k(t) = \sum_{j=1}^p \nu_{jk} \cdot f\left[\sum_{i=1}^m x \cdot \omega_{ij} + \theta_j\right] + r_t \\ f(x) = 1/(1 + \exp(-x)) \\ s.t. \quad w \in R^{m \times p}, v \in R^{p \times n}, \theta \in R^p, r \in R^n \end{array} \right. \quad (1)$$

where x is training samples, $\hat{y}_k(t)$, $y_k(t)$ are the desired output and real data, respectively. The fitness function is defined as follows:

$$F(w, v, \theta, r) = 1/(1 + \min E(w, v, \theta, r)) \quad (2)$$

Here we introduce our scheme:

Step 1: Initialize positions and speeds of a number of particles. M particles are randomly generated and each of them includes two parts: position and speed. The position of each particle consists of network node link and connection weights. The hidden nodes are encoded as binary code string, 1 with connection and 0 without connection. The connection weights are encoded as float string, randomly generated within $[-1, 1]$.

Step 2: Input training samples and calculate the fitness of each particle according to Expression (2). Initialize individual best position $P_{best}(t)$ and the global best position $P_{gbest}(t)$.

Step 3: Compare individual current fitness and the fitness of its experienced best position. If current fitness is better, we set current position to be the best position.

Step 4: Equation of speed evolution for each particle can be written as follows:

$$v_{ij}(t+1) = \omega(t) \cdot v_{ij}(t) + c_1 r_1 (P_{best}(t) - x_{ij}(t)). \quad (3)$$

$$\omega(t) = \omega_{max} - [(\omega_{max} - \omega_{min}) / (iter_{max})] \cdot iter. \quad (4)$$

where ω_{max} , ω_{min} denote the maximum and minimum of inertia weights, respectively. While $iter$, $iter_{max}$ denote current iteration number and the maximum iteration number.

Step 5: According to Ref. [14], equation of network link can be written as follows:

$$x_{ij}(t+1) = \begin{cases} 0, & r \geq 1/(1 + \exp(-v_{ij}(t))) \\ 0, & r < 1/(1 + \exp(-v_{ij}(t))) \end{cases} \quad (5)$$

where r ranges from $[0,1]$. Equation of position evolution for each particle can be written as follows:

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1). \quad (6)$$

Step 6: Repeat step 2 ~ 5 until stopping criteria are satisfied, e.g., the best fitness is satisfied or the maximum iteration number is reached.

Step 7: Decode each particle and obtain M groups network architecture and connection weights. Thus, we can form M different neural networks. Train these networks with training samples until stopping criteria are satisfied so that we will generate M different individual neural network predictors.

2.2 Selecting Appropriate Ensemble Members by the Principal Component Analysis (PCA)

After training, each individual neural predictor has generated its own result. However, if there are a great number of individual members, we need to select a subset of representatives in order to improve ensemble efficiency. In this study, the PCA technique [11] is adopted to select appropriate ensemble members. Interested readers can be referred to [11] for more details.

2.3 Combining the Selected Members by NR

The traditional linear and non-linear regression models fit the model

$$Y_i = f(\beta, X'_i) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (7)$$

where $\beta = (\beta_1, \beta_2, \dots, \beta_M)$ is a vector of parameters to be estimated, and $X'_i = (x_1, x_2, \dots, x_M)$ is a vector of predictors for the i th of n observations; the errors ε_i are assumed to be normally and independently distributed with mean 0 and constant variance σ^2 . The function $f(\cdot)$ relating the average value of the response y to the predictors, is specified in advance, as it is in a linear or nonlinear regression model [15].

The assumption that the pairs (X'_i, Y) , $i = 1, 2, \dots, n$ are an independent sample from an unknown distribution can often be justified in practice and simplifies technical matters for getting function $f(\cdot)$. Certainly in the expenditure data example in which the observations have been gathered from a quite realistic cross-section of the data, the assumptions of independence and identical distributions seem to be difficult [16]. Regression model realized that pure parametric

thinking in curve estimations often does not meet the need for edibility in data analysis and the development of hardware created the demand for theory of now computable nonparametric estimates.

The general nonparametric regression model is written in a similar manner, but the function $f(\cdot)$ is left unspecified:

$$Y_i = f(X_i') + \varepsilon_i, \quad i = 1, 2, \dots, n \tag{8}$$

Moreover, the object of nonparametric regression is to estimate the regression function $f(\cdot)$ directly, rather than to estimate parameters. Most methods of nonparametric regression implicitly assume that $f(\cdot)$ is a smooth, continuous function. As in nonlinear regression, it is standard to assume that $\varepsilon_i \sim N(0, \sigma^2)$. Nonparametric regression is a type of regression analysis in which the functional form of the relationship between the response variable and the associated predictor variables does not to be specified in order to fit a model to a set of data. This makes non-parametric regression a good competitor to non-linear regression for modelling situations in which a theoretical model is not known, or is difficult to fit. There are several approaches to estimating nonparametric regression models, for example Local Polynomial Regression, Smoothing- Spline regression, Local Likelihood regression, Kernel Function Regression, etc [17]. In this paper, Gauss Kernel Function Regression is defined as

$$Y_i = \frac{\sum_{i=1}^n K\left(\frac{x_i-x}{h_n}\right)Y_i}{\sum_{i=1}^n K\left(\frac{x_i-x}{h_n}\right)} + \varepsilon_i, \quad i = 1, 2, \dots, n \tag{9}$$

where $K(x) = (2\pi)^{0.5}exp(-0.5)x^2$, The condition $E(\varepsilon_i|Y_i) = 0$ is sufficient, then

$$E(\hat{Y}|Y) = \hat{m}(X) \tag{10}$$

$$\lim_{n \rightarrow \infty} P\{|\hat{m}(X) - m(X)| > \varepsilon\} = 0 \tag{11}$$

And h_n is called the bandwidth. Kernel function is a very important parameter, which can effectively eliminate random interference and smooth regression curve. The bandwidth is used to control regression model error, if the bandwidth value is too large, the regression model is linear; if the bandwidth value is too small, the regression function is not be smoothed and random interference is not eliminated. Theoretically, $h_n = cn^{-0.5}$ is best value, and C is constant, which is determined by cross-test [18].

2.4 NR Ensemble Rainfall Model

The above-mentioned method can be summed up as follows: firstly, PSO algorithm is applied to evolve neural network architecture and connection weights. The evolved neural network architecture and connection weights are input into a new neural network. The new neural network is trained using back-propagation (BP) algorithm, generating different individual neural network predictors. Secondly, the PCA technique extracts ensemble members. Finally, NR is used to combine the selected individual forecasting results into a ensemble model.

3 Selection of Data and Method of Model Building

3.1 Empirical Data

The Liu-Jian Watershed rainfall data is employed as a case study for development of rainfall forecasting model in this investigation. The watershed is located in southwest Guangxi of China, due to the southwest monsoon, monsoon trough and tropical cyclones, it is typically with heavy raining in the summer. Real-time ground rainfall data is obtained from May and June 1951 to May and June 2004 in Guangxi by observing 89 stations. Guangxi region has been divided into three regional precipitation by the group-average method. Statistics for each district in the average daily precipitation is used as the forecasting object. Figure.1 shows three region map.

In one district as an example to show the process of modelling, different climatic variability and its effect have been discussed many times in the literature. Based on routine weather materials and T213 numerical prediction product, firstly, the 500-hPa monthly mean geopotential height field of the Northern Hemisphere data, and the sea surface temperature anomalies in the Pacific Ocean data. We get 6 variables as the predictors by analyzing daily precipitation in one district. The original daily rainfall data is used as the predicted variables. Using rainfall data as training sample is builded modelling in May and June of 2005 and 2006, rainfall data is tested modelling in May and June of 2007. The training sample is 114 and testing sample is 55. In order to measure effectiveness of the proposed method, we compare results of NR-PSO-ANN model. Four types of errors are described as follows:

Maximum Absolute Error:

$$MAE_1 = \max\{|Y_i - \hat{Y}_i|, i = 1, 2, \dots, n\}. \quad (12)$$

Mean Absolute Error:

$$MAE_1 = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i|, i = 1, 2, \dots, n. \quad (13)$$

The Error Value more than 25mm:

$$F_1 = \sum_{i=1}^n I_i \quad (14)$$

$$\text{where } I_i = \begin{cases} 1, & |Y_i - \hat{Y}_i| > 25 \\ 0, & |Y_i - \hat{Y}_i| \leq 25. \end{cases}$$

The Error Value less than 5mm

$$F_1 = \sum_{i=1}^n L_i. \quad (15)$$

$$\text{where } L_i = \begin{cases} 1, & |Y_i - \hat{Y}_i| < 5 \\ 0, & |Y_i - \hat{Y}_i| \leq 5. \end{cases}, Y_i \text{ is original value, } \hat{Y}_i \text{ is forecast value.}$$

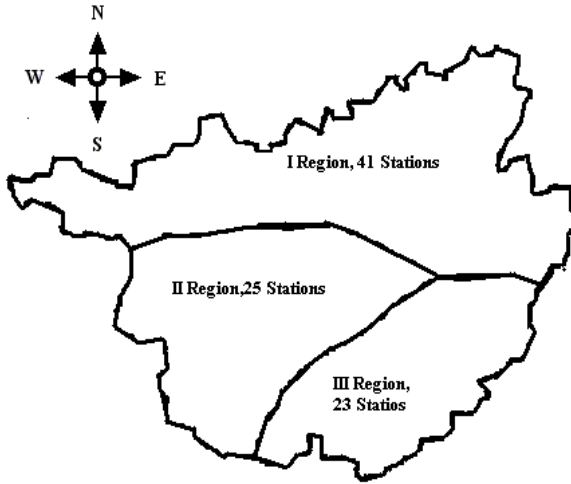


Fig. 1. The Group Average Region Map of Guangxi Rainfall

3.2 Analysis of the Results

PSO-BP parameters are set as follows: the iteration times are 100; the population is 100; the minimum inertia weight is 0.1; the maximum inertia weight is 0.9. BP parameters are set as follows: the number of neurons in the hidden layer range from 6 ~ 15. The learning rate is 0.9; the momentum factor is 0.7; the iteration times are 1000; the global error is 0.001.

When all the training results satisfy the request of error, all the component neural networks have been trained well. After training, the PCA approach is adopted to select some neural network output without linear correlation from the available neural networks to constitute an ensemble, and NR is used to combine their forecasting results. For the purpose of comparison, we have also built two other ensemble forecasting models: (1) stepwise linear regression (SLR) all the available forecasting output with feature extraction by PCA; (2) output of T213 numerical prediction product.

Figure 2 shows the curve of fitness in the training stage. One can see that the maximum, average and the minimum fitness and convergent speed are tending towards stability with increase of iteration number. Therefore, network architecture and connection weights are in near-optimal zone.

Each of the models described in the last section is estimated and validated by the same sample data. The model estimation selection process is then followed by an empirical evaluation based on the out-of-sample data. At this stage, the relative performance of the models is measured by four errors.

Figure 3 shows fitting of training sample, we can see that learning results of NR-PSO-ANN are satisfying. The more important factor to measure performance of a method is to check its generalization ability. Figure 4 and Figure 5 shows forecasting results. Table 1 reports the errors results. From figure 4,

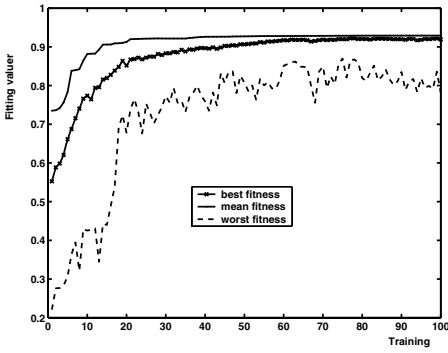


Fig. 2. Fitness Values in PSOBP Approach

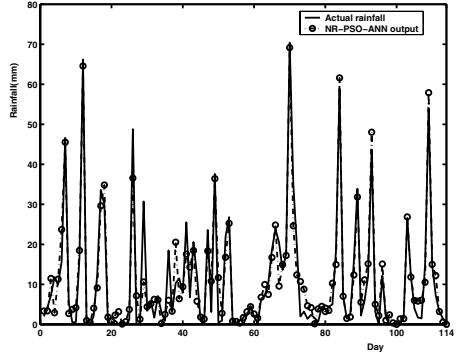


Fig. 3. Fitting of Training Samples

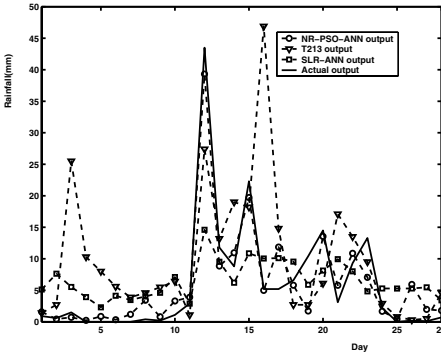


Fig. 4. Forecasting of Samples in May

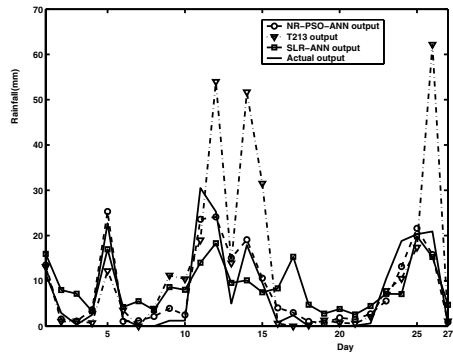


Fig. 5. Forecasting of Samples in June

figure 5 and Table 1, the differences between the different models are very significant. For example, for the rainfall test case in May, the MAE_1 for the T213 model is 41.30% and for the SLR-ANN model it is 16.50%, while for the proposed NR-PSO-ANN model, the MAE_1 reaches 10.04%, which has obvious advantages over other two models. The errors of the NR-PSO-ANN model has less than other model in June forecasting rainfall. The results imply that the NR-PSO-ANN has a significant forecasting ability under the same network input.

If the error value less than 5mm is reference information, the NR-PSO-ANN model has reference information at least 89%, which reference information of T213 is only 67%, and reference information of SLR-ANN is only 52% in May. If the error value more than 25mm is unreliable information, the NR-PSO-ANN model has no unreliable information, which unreliable information of T213 is 11% in May. The forecasting of NR-PSO-ANN model for a Meteorological application has greater forecasting accuracy.

Furthermore, we use the same method to train precipitation data and predict precipitation from May to June at the other two regions. The experimental

Table 1. A Comparison of Result of Ensemble Models about Test Samples

Errors	Month	NR-PSO	SLR	T213	Month	NR-PSO	SLR	T213
MAE_1	May	10.04	16.50	41.30	June	8.68	28.89	41.70
MAE_2		2.38	5.56	7.74		2.16	5.36	7.21
F_1		0	0	3		0	1	1
$F_1 \times 100\%$		0%	0%	11%		0%	4%	4%
F_2		24	14	18		24	18	16
$F_1 \times 100\%$		89%	52%	67%		86%	64%	57%

results also show that NR-PSO-ANN method is better generalization ability than SLR-ANN and T213 method.

4 Conclusions

In this paper, a novel NR ensemble approach, based on PSO algorithm and ANN, is represented for meteorological prediction, PCA technology combines linear characteristics with nonlinear characteristics, NR is used to combine the selected individual forecasting results into a ensemble model. The novel NR ensemble including nonparametric regression, PSO algorithm and ANN method outperform all linear ensemble methods, implying that the proposed nonlinear ensemble model can be used as a feasible approach to rainfall forecasting.

Acknowledgments

This work was supported in part by National Natural Science Foundation of China (No.60775037), Program for New Century Excellent Talents in University (No.NCET-05-0549) and in part by the Guangxi Science Foundation under Grant No. 0832092.

References

1. Hong, W.C.: Rainfall Forecasting by Technological Machine Learning Models. Applied Mathematics and Computation 200, 41–47 (2008)
2. Luk, K.C., Ball, J.E., Sharma, A.: An Application of Artificial Neural Networks for Rainfall Forecasting. Mathematical and Computer Modelling 33, 683–693 (2001)
3. Nasser, M., Asghari, K., Abedini, M.J.: Optimized Scenario for Rainfall Forecasting Using Genetic Algorithm Coupled with Artificial Neural Network. Expert Systems with Application 35, 1414–1421 (2008)
4. Hsieh, W., Tang, B.: Applying Neural Network Models to Prediction and Data Analysis in Meteorology and Oceanography. Bull. Am. Meteorol. Soc. 79, 1855–1870 (1998)
5. Valverde, M.C., Campos Velho, H.F., Ferreira, N.J.: Artificial Neural Network Technique for Rainfall Forecasting Applied to the Sö Paulo Region. Journal of Hydrology 301(1-4), 146–162 (2005)

6. Wang, W., Xu, Z., Lu, J.W.: Three Improved Neural Network Models for Air Quality Forecasting. *Engineering Computations* 20(2), 192–210 (2003)
7. Hykin, S.: *Neural Networks: A Comprehensive Foundation*. Printice-Hall, Inc., New Jersey (1999)
8. Lin, G.F., Chen, L.H.: Application of an Artificial Neural Network to Typhoon Rainfall Forecasting. *Hydrological Processes* 19, 1825–1837 (2005)
9. Wu, J.S., Jin, L.: Forecast Research and Applying of BP Neural Network Based on Genetic Algorithms. *Mathematics in Practice and Theory* 35(1), 83–88 (2005)
10. Wu, J.S., Jin, L., Liu, M.Z.: Modeling Meteorological Prediction Using Particle Swarm Optimization and Neural Network Ensemble. In: Wang, J., Yi, Z., Zurada, J.M., Lu, B.-L., Yin, H. (eds.) *ISNN 2006*. LNCS, vol. 3973, pp. 1202–1209. Springer, Heidelberg (2006)
11. Yu, L., Wang, S., Lai, K.K.: A Novel Nonlinear Ensemble Forecasting Model Incorporating Glar Andann for Foreign Exchange Rates. *Computers Operations Research* 32, 2523–2541 (2005)
12. Brandstatter, B., Baumgartner, U.: Particle Swarm Optimization-Mass-Spring System Analogon. *IEEE Transactions on Magnetics* 38, 997–1000 (2002)
13. Fan, S.K., Liang, Y.C.: Hybrid Simplex Search and Particle Swarm Optimization for the Global Optimization of Multimodal Functions. *Engineering Optimization* 36, 401–418 (2003)
14. Kennedy, J., Spears, W.: Matching Algorithms to Problems: an Experimental Test of the Particle Swarm and Some Genetic Algorithms on the Multimode Problem Generator. In: *IEEE International Conference on Evolutionary Computation*, Anchorage, Alaska, USA (1998)
15. Nason, G.P., Silverman, B.W.: Wavelets for Regression and Other Statistical Problems. In: Schimek, M.G. (ed.) *Smoothing and Regression: Approaches, Computation, and Application*. Wiley, New York (2000)
16. Fox, J.: *Multiple and Generalized Nonparametric Regression*. Sage, Thousand Oaks (2000)
17. Ioannides, D.A., Alevizos, P.D.: Nonparametric Regression with Errors in Variables and Application. *Statistics & Probability Letters* 32, 35–43 (1997)
18. Shen, H.P., Lawrence, D.B.: Nonparametric Modelling of Time-Varying Customer Service Times at Bank Call Centre. *Applied Stochastic Models in Business and Industry* 22, 297–311 (2006)

A Revised Neural Network for Solving Quadratic Programming Problems

Yinjie Sun

Henan Normal University, 453007, Xinxiang, Henan, China
sunyinjie@126.com

Abstract. By selecting an appropriate transformation of the variables in quadratic programming problems with equality constraints, a lower order recurrent neural network for solving higher quadratic programming is presented. The proposed recurrent neural network is globally exponential stability and converges to the optimal solutions of the higher quadratic programming. An op-amp based on the analogue circuit realization of the recurrent neural network is described. The recurrent neural network proposed in the paper is simple in structure, and is more stable and more accuracy for solving the higher quadratic programming than some existed conclusions, especially for the case that the number of decision variables is close to the number of the constraints. An illustrative example is discussed to show us how to design the analogue neural network using the steps proposed in this paper.

Key words: Quadratic Programming Problems, Neural Network, Analogue Circuit.

1 Introduction

Quadratic programming problems are very important in the field of optimization. They arise in many applications such as constrained least mean square estimation. Besides its wide applications, quadratic programming is also of theoretic meaning, because it forms a basis for solving some general nonlinear programming problems. Many effective algorithms on quadratic programming have been developed [1], such as the elimination method, orthogonal decomposition method and the Lagrangian multiple method. As we all know, the programs programmed by these algorithms execute in serial way, not parallel form. In many practical applications, it is desired that the quadratic programming problems be solved in real time. In such applications, the effective parallel procedures are required.

Neural networks have been exhibited the abilities of parallel and distributed computation, they fit for solving real time problems. In recent years, neural networks have been used for solving optimization problems widely, see References [2,3,4,5,6,7]. Two approaches are used to design recurrent neural networks for optimization in these papers. One is based on a defined energy function and the other is based on the existing optimal conditions, such as Kuhn-Tucher necessary condition. Once the recurrent neural networks are well designed, the stability

analysis of the neural networks must be done [8,9] and the realized circuits of the neural networks should be discussed.

In quadratic programming problems, the objective function is quadratic and the constraints are linear. Let x be an n -dimensional column vector of decision variables, i.e. $x \in \mathbb{R}^n$, a quadratic programming problem with equality constraints can be described as follows:

$$\begin{aligned} \text{minimize} \quad & q(x) = \frac{1}{2}x^T G x + g^T x, \\ \text{subject to} \quad & A^T x = b. \end{aligned} \quad (1)$$

Where G is an $n \times n$ matrix of objective coefficients, i.e. $G \in \mathbb{R}^{n \times n}$. We assume the matrix G be symmetrical but not be positive definite. $g \in \mathbb{R}^n$ is the column vector of objective coefficients. $A \in \mathbb{R}^{n \times m}$ ($m < n$), each column of A is formed by the coefficients of the corresponding constraint. We assume that the matrix A be full rank. $b \in \mathbb{R}^m$ is the vector of constraint coefficients. The superscript T denotes the transpose operator.

For (1), we will select an appropriate transformation of the variables, and based on it, an $(n - m)$ -dimensional recurrent neural network for solving (1) is presented. The proposed recurrent neural network is globally exponential stability and converges to the optimal solutions of (1). An op-amp based on the analogue circuit realization of the recurrent neural network is described. The recurrent neural network proposed in the paper is quicker and more accuracy to solve the higher quadratic programming (1) than some existed conclusions, such as in [5], especially for the case that the number of decision variables n is close to the number of the constraints m , because in this case $n - m$ is a small integer. An illustrative example is discussed to show the effectiveness of the analogue neural network.

2 Linear Transformation of Variables

Select two matrices $S \in \mathbb{R}^{n \times m}$ and $Z \in \mathbb{R}^{n \times (n-m)}$, such that $A^T S = I_m$, $A^T Z = 0$ and $(S : Z) \in \mathbb{R}^{n \times n}$ is nonsingular, I_m is m -dimensional unit matrix. That is, the matrix S is the pseudo-inverse of the matrix A , and A is column full rank, so $S = A(A^T A)^{-1}$. The n -dimensional column vectors z_1, z_2, \dots, z_{n-m} of the matrix Z is the base vector group of the linear space $N(A) = \{\delta \in \mathbb{R}^n : A^T \delta = 0\}$. For any $\delta \in N(A)$, there exist an unique vector $y = (y_1, y_2, \dots, y_{n-m})^T \in \mathbb{R}^{n-m}$, such that

$$\delta = \sum_{i=1}^{n-m} y_i z_i = Z y.$$

For any feasible point $x \in \{x \in \mathbb{R}^n : A^T x = b\}$, we have $A^T x = b$, the constraint equation has a particular solution Sb , and its general solutions have the form: $x = Sb + \delta$, $\delta \in N(A)$. So, we choose the linear transformation:

$$x = Z y + Sb, \quad y \in \mathbb{R}^{n-m}. \quad (2)$$

Transformation (2) constructs a one to one mapping between the constraint set $\{x \in \mathbb{R}^n : A^T x = b\}$ and \mathbb{R}^{n-m} and provides an approach using an $(n-m)$ -dimensional vector y to eliminate the constraint $A^T x = b$.

Replacing x in (1) by (2), we have the following lower order unconstrained optimization problems:

$$\text{minimize } q(x) = \phi(y) = \frac{1}{2}y^T(Z^T GZ)y + (g + GSb)^T Zy + \frac{1}{2}(2g + GSb)^T Sb. \quad (3)$$

Lemma 1. x^* is the global minimum point of (1) if and only if there exists a matrix Z , such that the matrix $Z^T GZ$ is positive definite.

Proof. Noticing the two facts: 1. For the unconstrained quadratic programming problem (3), $y^* \in \mathbb{R}^{n-m}$ is the global minimum point if and only if its Hessian matrix $Z^T GZ$ is positive definite; 2. Transformation (2) is a one to one mapping.

Sufficiency: Assume the matrix $Z^T GZ$ is positive definite, then, (3) has an unique global minimum point y^* , thus $x^* = Zy^* + Sb$ is the global minimum point of (1).

Necessity: Assume x^* is the global minimum point of (1), then there exists an unique y^* , such that $x^* = Zy^* + Sb$ and y^* is the unique global minimum point of (3), thus the matrix $Z^T GZ$ is positive definite. \square

Remark 1. Obviously, if the Hessian matrix G in (1) is positive definite, then $Z^T GZ$ is positive definite, but the inverse proposition is not true; If we can determine (1) has the unique minimum from the practical sense, then, there exists a matrix Z such that $Z^T GZ$ is positive definite.

In our paper, we always assume that the quadratic programming problem (1) has an unique global minimum point or equivalently, there exists the matrix Z , such that the matrix $Z^T GZ$ is positive definite, not assume the matrix G is positive definite.

Remark 2. The existence of the matrices S, Z can be illustrated by the orthogonal decomposition method. The QR decomposition of the matrix A has the form:

$$A = Q \begin{pmatrix} R \\ 0 \end{pmatrix} = (Q_1 \quad Q_2) \begin{pmatrix} R \\ 0 \end{pmatrix} = Q_1 R.$$

Where, $Q \in \mathbb{R}^{n \times n}$ is an orthogonal matrix, $R = (r_{ij}) \in \mathbb{R}^{m \times m}$ is an upper triangular matrix with $r_{ii} > 0$, $Q_1 \in \mathbb{R}^{n \times m}$ and $Q_2 \in \mathbb{R}^{n \times (n-m)}$. It is easy to verify that

$$S = Q_1 R^{-T}, \quad Z = Q_2$$

satisfy the properties, i.e. $A^T S = I_m$, $A^T Z = 0$ and $(S : Z)$ is nonsingular.

3 Neural Network Design

For (3), by setting the gradients of $\phi(y)$ to zero, the necessary condition gives rise to the following matrix form algebraic equation:

$$(Z^T GZ) y = -Z^T (g + GSb). \quad (4)$$

In order to get the real time solution of (4), it can be realized by the analogue circuits of the recurrent neural network. The dynamical equations of the recurrent neural networks is as follows:

$$\frac{dy}{dt} = \alpha W y + \alpha \theta. \quad (5)$$

Where $\alpha > 0$ is a scaling adjustable parameter, $W = -Z^T G Z$ is the $(n - m)$ -dimensional connection weight matrix, $\theta = -Z^T (g + G S b)$ is the $(n - m)$ -dimensional biasing threshold vector.

Theorem 1. *Assume that the quadratic programming problem (1) has an unique global minimum point, then the dynamical equation (5) has an unique equilibrium y^* , and y^* is global exponential stability and the convergent rate is $-\alpha \lambda_{\max}(W) > 0$, where $\lambda_{\max}(W)$ denotes the maximum eigenvalue of the matrix W .*

Proof. By the assumption of this theorem and Lemma 1, the matrix W is negative definite, all eigenvalues of the matrix W are negative. From the linear system theory, (5) has an unique equilibrium y^* , and for the solution $y(t)$ of (5) has the estimation as follows :

$$\|y(t) - y^*\| \leq \|y(0)\| e^{\alpha \cdot \lambda_{\max}(W) \cdot t}.$$

Where $y(0)$ is the initial value of (5). Thus, y^* is global exponential stability and the convergent rate is $-\alpha \lambda_{\max}(W) > 0$. \square

Remark 3. The convergent speed of the equilibrium y^* of the recurrent neural network (5) can be adjustable by the parameter α , more larger the value of α , more shorter the transient processes.

Once the equilibrium y^* is obtained, the optimal solution of the primal quadratic programming problem (1) can be computed by $x^* := Z y^* + S b$.

Fig. 1 is an op-amp based on analogue circuit scheme diagram of the neural network for solving quadratic programming problem with equality constrains (1), it consists of two parts.

The lower part of Fig. 1 is the circuit realization of the recurrent neural network (5). It contains $n - m$ neurons y_1, \dots, y_{n-m} , each neuron can be realized by a summer, an integrator, and an inverter. The ohmic value of each connection resistor is determined according to the magnitude of the corresponding connection weight, i.e. $R_{ij} = R_f / |w_{ij}|$, $i, j = 1, 2, \dots, n - m$. The Connecting terminal of each connection resistor is determined according to the sigh of the connection weight; i.e. if the connection from neuron j to neuron i is excitatory (i.e. $w_{ij} > 0$), then connect R_{ij} to the terminal y_j ; If the connection from neuron j to neuron i is inhibitory (i.e. $w_{ij} < 0$), then connect R_{ij} to the terminal $-y_j$. The biasing threshold for neuron i can be realized by a voltage source E_i such that $\theta_i = R_f E_i / R_i$, $i = 1, \dots, n - m$; The adjustable parameter α can be realized by the circuit time constant, i.e. $\alpha = 1 / (R_c C)$.

The upper part of Fig. 1 realized the algebraic equation (2) which is a forward neural network for solving the optimal solution of quadratic programming

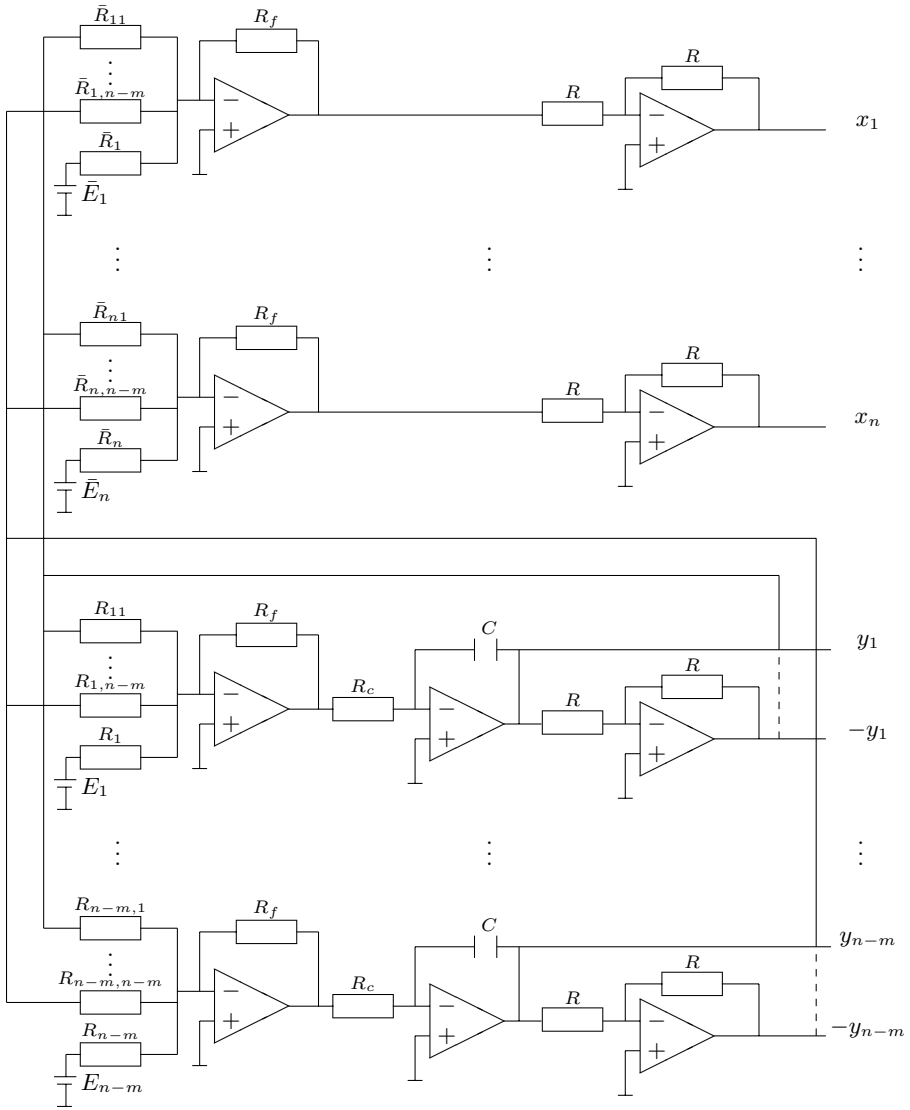


Fig. 1. Op-amp based on circuit realization of analogue neural network

problem (II). It contains n neurons x_1, \dots, x_n , each neuron is composed of a summer and an inverter, no integrator. The ohmic value of each connection resistor is determined by the connection weight matrix Z , i.e. $\bar{R}_{ij} = R_f/|z_{ij}|, i = 1, \dots, n, j = 1, \dots, n - m$, and \bar{R}_{ij} connects terminal y_j if $z_{ij} > 0$ or connects $-y_j$ if $z_{ij} < 0$. The biasing threshold vector Sb can be realized by the voltage sources $\bar{E}_i, i = 1, \dots, n$, i.e. $Sb = (R_f \bar{E}_1 / \bar{R}_1, \dots, R_f \bar{E}_n / \bar{R}_n)^T$.

Remark 4. 1. The total neurons in Fig. 1 are $(n-m)+n = 2n-m$. If $n \approx m$, i.e. the number of decision variables n and the number of constraints m of (1) are almost equal, the total neurons in Fig. 1 are close to n . In Reference [5], it needs $n+m \approx 2n$ neurons. For this case, we see that the neural network proposed in this paper needs half the neurons in the analogue circuit of [5];

2. The number of feedback neurons (the lower part) in Fig. 1 is $n-m$, if $n \approx m$, we see that the dimension of the current neural network is very low. So, the analogue circuit realization of neural network solving the quadratic programming problem (1) is simple in structure, and is more stable and more accurate, the transient processes is shorter than some existed neural networks.

4 An Illustrative Example

The following illustrative example shows us how to design the analogue neural network using the steps proposed in this paper.

Consider the following numeric example with the following coefficients [5]:

$$G = \begin{pmatrix} 3 & -1 & 1 & -2 \\ -1 & 4 & 2 & 0 \\ 1 & 2 & 5 & 1 \\ -2 & 0 & 1 & 6 \end{pmatrix}, g = \begin{pmatrix} -6 \\ 15 \\ 9 \\ 4 \end{pmatrix}, A^T = \begin{pmatrix} 1 & 2 & 4 & 5 \\ 3 & 2 & 1 & -2 \end{pmatrix}, b = \begin{pmatrix} 12 \\ -9 \end{pmatrix}.$$

The orthogonal-triangular decomposition of the matrix A is $A = Q(R^T, 0)^T$ (use the MATLAB function 'qr'), where

$$Q = \begin{pmatrix} -0.1474 & -0.7024 & \vdots & -0.5434 & -0.4354 \\ -0.2949 & -0.4614 & \vdots & 0.0060 & 0.8367 \\ -0.5898 & -0.2153 & \vdots & 0.7041 & -0.3317 \\ -0.7372 & 0.4973 & \vdots & -0.4570 & 0.0177 \end{pmatrix} = (Q_1 : Q_2),$$

$$R = \begin{pmatrix} -6.7823 & -0.1474 \\ 0 & -4.2401 \end{pmatrix}.$$

Choose the matrices S and Z , such that $S = Q_1 R^{-T}$, $Z = Q_2$. We have

$$Z^T G Z = \begin{pmatrix} 2.2392 & 0.8041 \\ 0.8041 & 3.8474 \end{pmatrix}, \bar{\theta} := S b = \begin{pmatrix} -1.2733 \\ -0.4860 \\ 0.5733 \\ 2.3905 \end{pmatrix}.$$

Because the matrix $Z^T G Z$ is positive definite, the optimal solution of this quadratic programming problem has an unique global minimum (use Lemma 1). In the next, we design the analogue neural network (5) and (2). The connection weight matrix and the biasing threshold vector are as follows:

$$W = \begin{pmatrix} -2.2392 & -0.8041 \\ -0.8041 & -3.8474 \end{pmatrix}, \theta = \begin{pmatrix} -6.1003 \\ -15.2404 \end{pmatrix}.$$

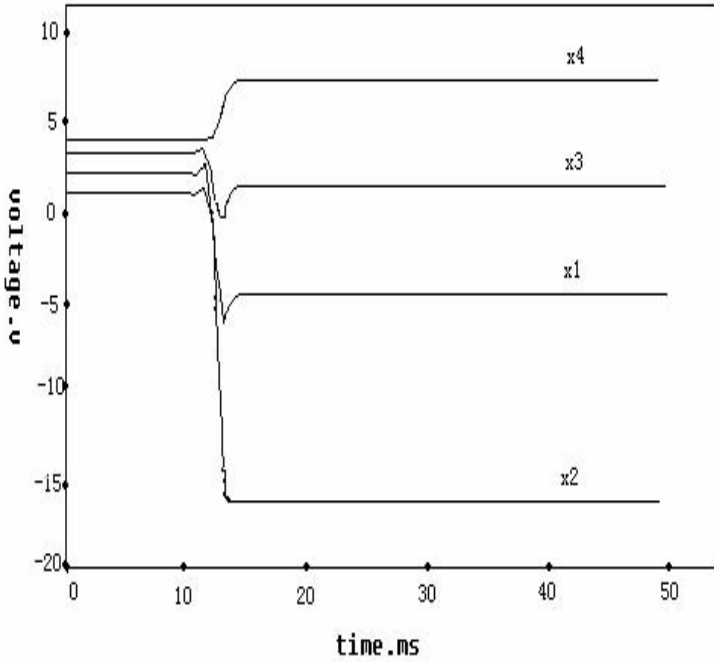


Fig. 2. The solutions of the quadratic programming problem simulated by EWB

Let $R_f = 5\text{k}\Omega$, $R_i = 1\text{k}\Omega$, $\bar{R}_i = 5\text{k}\Omega$, $R_c = 1\text{k}\Omega$, $C = 1\mu\text{F}$. The connection resistance matrix in kilo-ohms and voltage array in volts can be determined as follows, where $R_{ij} = R_f/w_{ij}$, $E_i = (R_i/R_f)\theta_i$, $\bar{R}_{ij} = R_f/z_{ij}$, $\bar{E}_i = (\bar{R}_i/R_f)\bar{\theta}_i$.

$$[R_{ij}] = \begin{pmatrix} -2.2329 & -6.2179 \\ -6.2179 & -1.2996 \end{pmatrix}, [E_i] = \begin{pmatrix} -1.2201 \\ -3.0481 \end{pmatrix},$$

$$[\bar{R}_{ij}] = \begin{pmatrix} -9.2013 & -11.4837 \\ 833.3333 & 5.9759 \\ 7.1013 & -15.0739 \\ -10.9409 & 282.4859 \end{pmatrix}, [\bar{E}_i] = \begin{pmatrix} -1.2733 \\ -0.4860 \\ 0.5733 \\ 2.3905 \end{pmatrix}.$$

According to these magnitudes of the resistances and the voltage sources in Fig. 1, we can design the analogue circuit and simulate by the electronic simulation software EWB to get the solutions of the quadratic programming problem with equality constraints, see Fig 2.

References

1. Fletcher, M.: Practical Methods of Optimization. John Wiley & Sons, Chichester (1981)
2. Hopfield, J.J., Tank, D.W.: Neural Computation of Decisions in Optimal Problems. Biological Cybernetics 52(3), 141–152 (1985)

3. Kennedy, M., Chua, L.O.: Neural Networks for Nonlinear Programming. *IEEE Trans.*, CAS-35(5), 554–562 (1988)
4. Cheng, L. Hou, Z.-G., Tan, M., Wang, X.: A simplified recurrent neural network for solving nonlinear variational inequalities. In: *Proceedings of International Joint Conference on Neural Networks*, pp. 104–109 (2008)
5. Wang, J.: Recurrent Neural Network for Solving Quadratic Programming Problems with Equality Constraints. *Electronics Letter* 28(14), 1345–1347 (1992)
6. Wang, J.: Primal and Dual Neural Networks for Shortest-path Routing. *IEEE Transactions on Systems, Man and Cybernetics-Part A: Systems and Humans* 28(6), 864–869 (1998)
7. Xia, Y., Wang, J.: Recurrent Neural Networks for Solving Nonlinear Convex Programs with Linear Constraints. *IEEE Transactions on Neural Networks* 16(2), 379–386 (2005)
8. Liao, W., Wang, D., Wang, Z., Liao, X.: Stability of Stochastic Cellular Neural Networks. *Journal of Huazhong Univ. of Sci. and Tech.* 35(1), 32–34 (2007)
9. Liao, W., Liao, X., Shen, Y.: Robust Stability of Time-delyed Interval CNN in Noisy Environment. *Acta Automatica Sinica* 30(2), 300–305 (2004)

The Separation Property Enhancement of Liquid State Machine by Particle Swarm Optimization

Jiangshuai Huang, Yongji Wang, and Jian Huang

Key Lab.for Image Processing &Intelligent Control,
Department of Control Science and Engineering,
Huazhong University of Science and Technology,
Wuhan, 430074 China
wangyjch@hust.edu.cn

Abstract. The separation property of Liquid State Machine (LSM) is a key for its power of computing, but the weights and delays of the inter-connections in the spiking neural circuit are usually randomly created and kept unchanged, which hinders the performance of the LSM greatly. In this paper, particle swarm optimization (PSO) was applied to optimize the weights and delays of the circuit so as to enhance the separation property of the LSM. Separation of random spike trains and *Fisheriris data-set* classification experiments are done by the optimized circuit. Demonstration examples show that the PSO can enlarge the separation property of the circuit greatly compared to the normal Hebbian-learning algorithm and enhance the computing ability of LSM.

Keywords: Liquid state machine, Particle swarm optimization, Separation property.

1 Introduction

The recurrent neural network (RNN) has been exploited extensively because of its good nature. Compared to the feedforward neural network, the loops in the RNN can hold the input signals for a while so the RNN can retain the contextual relationship of the input signals. That is a key for dealing with temporal signals. However, the training for the RNN is not so easy. Atiya [1] introduced some training algorithms of RNN like the back-propagation through time but they are not simple as the back-propagation algorithm in feed-forward neural networks, which hinders the RNN from been applied extensively. In order to exploit the power of the RNN, a new computing model called a neural microcircuit or reservoir computing was developed. This kind of model can be divided to three parts structurally: the input layer, the circuit and the readout layer. The circuit is a kind of recurrent neural network and it is created stochastically but whose weights are never being trained during the computing process. In the readout layer, some simple algorithms like linear regression are adopted to update weights of this layer. Simple as it seemed, however it has achieved great computing ability and it has been applied to many problems [2]. Mainly there are three kinds of computing model of neural microcircuit. They are the Echo State Network [3], the Liquid State Machine [4] and the Backpropagation-Decorrelation [5].

In Liquid state Machine, the neurons in the microcircuit are spiking neurons [6], which deal with the spike directly. So the Liquid State Machine (LSM) processes the signals which are in form of spike trains. There are a lot of signals that are in this form, for example, cortical signals from the cortex. While traditionally those spike train like the cortical signals have to be counted into numerical series [7] before being processed.

The neurons in the circuit are typically of LIF model [4] and they are arranged in a 3D cube, see Fig.1. In LSM, the connection of two neurons is formed under a probability. We randomly choose 20% neurons to be the inhibitory neurons and force them to give negative output. So all the weights are positive. The connection probability of two neurons in the circuit is as following equation:

$$p_{(i,j)} = c \cdot \exp(-D_{(i,j)} / \lambda^2) \quad (1)$$

where $D_{(i,j)}$ is the distance of neuron i, j in liquid circuit; $c=1$ and $\lambda=2.23$. When the input signals are fed into the circuit, they will run in the circuit for a while before they converge into the stable state. Just as Maass et al. mentioned [4] that the LSM's computing mechanism was based on the perturbation, which is the disturbed potential of the neurons in the circuit, not the stable state. When input signals $u(t)$ are fed into the circuit, they run in the circuit and generate new spikes before the system converges to a stable state. The perturbation of each neuron in the circuit can be collected as a vector, which is called the liquid state $x^M(t)$ [4]:

$$x^M(t) = (L^M u)(t) \quad (2)$$

The L^M could be seen as a mechanism to collect the liquid state $x^M(t)$. In this paper, we record the membrane potentials to orderly vectors and make them the liquid state. After that, the liquid states are fed to the readout layer as input signals and get an output signals $y(t)$.

$$y(t) = f^M(x^M(t)) \quad (3)$$

Briefly speaking, the f^M is the algorithm adopted by the readout layer. In this paper we adopt the supervised neural network. Intuitively, we know that the different inputs form different liquid states in the circuit, and these liquid states are different from each other with certain distances in a certain measurement which will be stated later. The distances affect the classification precision greatly since the algorithm adopted in the readout layer is quite simple usually. So the separation property [4] of the LSM, which means the ability to separate different input patterns in the liquid circuit, is very important for the LSM's performance. Usually the circuit is created in a stochastic way and kept fixed. If we can use some approaches to update the weights of LSM, the separation property of the LSM may be increased, and the classification ability will be then enhanced.

2 Separation Property Definition

The diagram of Liquid State Machine is shown in figure 1. In which the classification process is roughly described. After the input spike trains are fed into input layer, they

are mapped into liquid circuit, forming liquid state vector, then classification was done through the readout layer. From figure 1 we can see that the separation of inputs is very important.

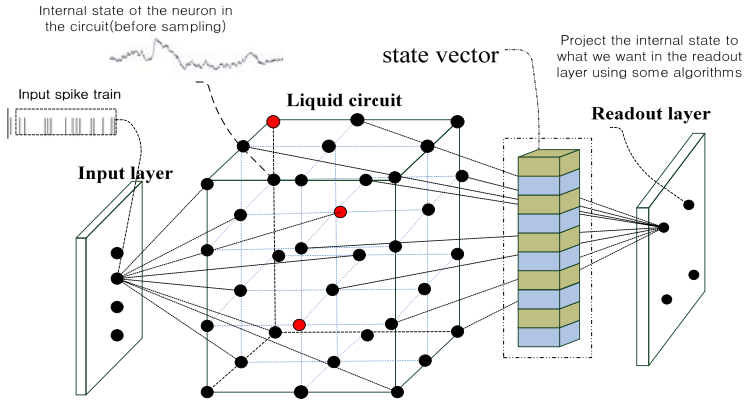


Fig. 1. Diagram of Liquid State Machine, the spiking neurons are arranged in a suppositional 3D cube. The red ones denote the inhibitory spiking neurons and the blacks denote the excitatory ones. The state vector is the snapshot of the membrane potential of the neurons.

Separation refers to “the amount of separation between trajectories of internal states of the system that are caused by two different input streams” [4]. Suppose that internal states O are the snapshot vectors of the neuron potentials of the same class, the L_2 norm can stand for the amount of separation. Here is the definition [8]:

$$Sep(\psi, O) = \frac{\sum_{i=1}^N \sum_{j=1}^N \|C_m(O_i(t)) - C_m(O_j(t))\|_2}{N^2} \quad (4)$$

where, C_m is the center of mass for each class of vectors; N is the total number of classes and $\|\cdot\|$ is the L_2 norm. The O_i, O_j are centers of two different trajectories caused by two input streams. $\|C_m(O_i(t)) - C_m(O_j(t))\|_2$ is a value that scales the distance of the two trajectories.

3 Particle Swarm Optimization (PSO)

In order to increase the separation ability of the circuit effectively, the particle swarm optimization (PSO) [9] was adopted in which the sets of weights and delays among the synapses in the circuit are treated as particles. PSO is a meta-heuristic approach motivated by the observation of the social behaviour of composed organisms, such as bird and fish flocks. Its main concept is that the knowledge which drives the search for the optimization is amplified by the social cooperation. The execution agents which perform the social interaction are called particles. The primary strategy of the

PSO is that each particle keeps track of its coordinates in an N-dimensional problem space which is related to the optimal solution they are seeking. Initially, the PSO generates a swarm of particles and each particle represents a possible solution to the problem and they update themselves during iterations based on the fitness function and the cooperation. During each iteration, the accelerating direction of one particle is determined by its own best solution found so far and the global best solution discovered so far by any of the particles in the swarm. This means that if a particle discovers a promising new solution, all the other particles will move closer to it, exploring the region more thoroughly in the process.

Let each particles have a current position in search space X_i , a current velocity V_i and a personal best position X_{pbest_i} and donate the global best position by X_{gbest} . In every step, the particles update their velocities and positions according to equation (5) and (6), respectively,

$$V_i(t+1) = wV_i(t) + c_1r_1(X_{pbest_i}(t) - X_i(t)) + c_2r_2(X_{gbest}(t) - X_i(t)) \quad (5)$$

$$X_i(t+1) = X_i(t) + V_i(t+1) \quad (6)$$

```

Initialize the population
for generation = 1 to max
  for i = 1 to swarmsize
    if  $f(x_i) < f(x_{pbest_i})$ 
       $x_{pbest_i} = x_i$ ;
    end
  end
   $j = \arg \min_i (x_{pbest_i})$ ;
   $x_{gbest} = x_{pbest_j}$ ;
  for  $d = 1$  to swarmsize
     $V_i(t+1) = wV_i(t) + c_1r_1(x_{pbest_i}(t) - x_i(t))$ 
       $+ c_2r_2(x_{gbest}(t) - x_i(t))$ ;
     $x_i(t+1) = x_i(t) + V_i(t+1)$ ;
  end
end

```

Fig. 2. Pseudocode of PSO

where, parameter c_1 and c_2 are set to be constant value, usually taken as 2, r_1 and r_2 are two random values distributed in $[0,1]$ and stay steady in the whole process. w is an inertia weight which controls the influence of previous velocity on the new velocity and it is set to 0.9. The pseudo-code of PSO is shown in Fig.2.

In this paper, the weights and delays of the synapses are treated as the particles and they are updated in parallel. There are 125 neurons in a circuit so a 125×125 vector can describe the weights or delays of the circuit explicitly. A particle is a set of two

125×125 vectors which records the weights and delays of the inter-connections among the neurons in the circuit separately. The initial particles are set to be positive values and if there are any negative values shows up we force them to be zero because the excitory neurons won't give inhibitory output and neurons can't take future input.

4 Demonstration Experiment Results

Two experiments are carried out, the exploring of the separation property's change after the training with PSO and the standard classification of *fisheriris data-set* before and after training with PSO. The same experiment was done with Hebbian-learning in order to compare. The basic experimental result was done by Hebbian-learning is in paper [10].

4.1 Separation Property

The whole process of this experiment is the training of the weights and delays of the synapses in the circuit with PSO. The size of the swarm is chosen as 30 and the number of iteration is about 50. The weights and delays are updated in parallel. After each iteration, we calculate the amount of separation according to equation (4). The separation is based on two liquid states and the liquid states are the mapping of two randomly generated spike-trains. After we feed the LSM with these two spike trains separately, there are two liquid states generated by these two spike trains. On enhancing the distance of these two liquid states, we use the PSO to update the weights and delays of the circuit. The LSM is such a kind of machine that project the input streams into a higher dimensional space, which make the classification easier. The Separation Property of LSM measures the simplicity of classification. Any tiny difference between the input streams may be magnified by the circuit. And the PSO in this paper will help enhancing the magnification no doubt. The whole process is done in the same way with Hebbian-learning. In Hebbian-learning, the synapse's weight changes in proportion to the temporal correlation between the presynaptic neurons and postsynaptic neurons. If the presynaptic neuron fires first, the weight will increase; otherwise the weight will decrease. So it only changes the weights of the circuit. The result is shown in Fig. 3.

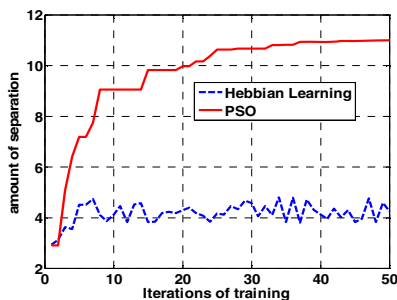


Fig. 3. Separation result of the PSO and Hebbian-learning given two randomly generated spike-trains

It is obvious that PSO can enlarge the separation property greatly while the Hebbian-learning has only small effect on the separation property. Compared with the Hebbian-learning, PSO have spent more time on the training to get the optimal weights and delays, Fig. 4 shows the spent time of two training algorithms.

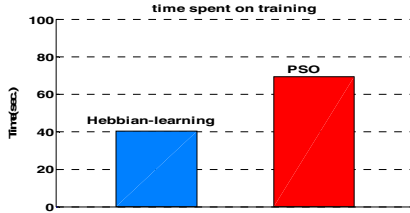


Fig. 4. comparison of time spent on two training algorithms. Comparatively the PSO doesn't spend too much time.

The changes of synapses' weights in the neural circuit before and after training are shown in Fig. 5. In that case there are 125 neurons arranged in a $5 \times 5 \times 5$'s cube. The weights were initialized according to the probability of equation (1). From Fig. 5 we can see that the training process of PSO is embodied on the gradually changing of the synapses weights. The colour of the lines denotes the weight strength. The darker the colour is, the weaker the weight is. After number of training iterations, many synapses are weakened strongly while the remainders are strengthened.

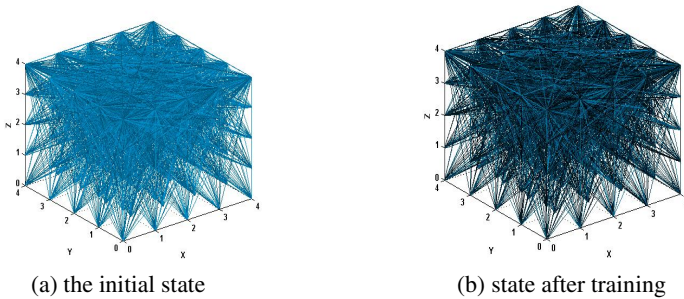


Fig. 5. The change of weights. The colour of the lines denotes the weight strength. The darker the colour is, the weaker the weight is.

The initial weights and delays randomly generated and unchanged in neural circuits will result in two kinds of weakness frequently, which are the *Pathological Synchrony* and *Over-Stratification* (see Fig.6 and Fig.7). The *Pathological Synchrony* happens when all neurons are in positive feedback circles and they fire and arose others to fire. This behaviour won't help separating because every pattern gets the same responses in the circuit. The *Over-Stratification* is opposite to the above one. Every neuron gets silent and can't evoke others. Few of neurons fire just when the spikes from input layer come.

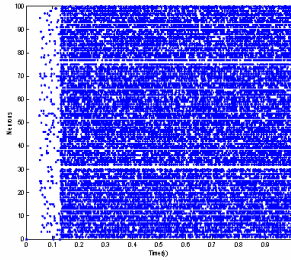


Fig. 6. Pathological Synchrony. This may happen when neurons in the circuit get excitory feedbacks from each other and cause to fire infinitely. The dots represent the spikes.

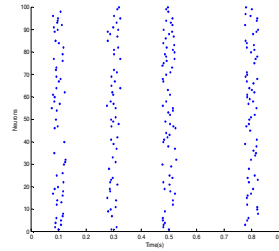
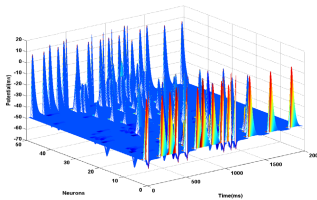


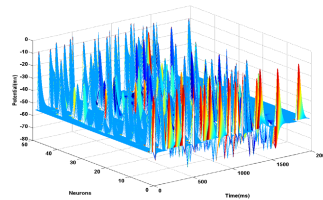
Fig. 7. Over-Stratification. Opposite to the Pathological Synchrony, the Over-Stratification appears when the silence haunts the circuit; the neurons can't evoke each other.

Because the PSO can optimize weights and delays of the circuits, it is able to improve these morbidities. Along training, the *Pathological Synchrony* and *Over-Stratification* are decreasing gradually. Fig.8 shows the melioration of *Over-Stratification*, by showing variation over time of the membrane potentials of all neurons, before and after training of random input. In Fig. 8(a) most neurons' potentials have no variation except for which receive the input spikes but in fig. 8(b) all neurons are active after training.

Fig. 9 shows the melioration of *Pathological Synchrony*. The initial state shows most neurons fire constantly but after 50 training iterations, the *Pathological Synchrony* doesn't exist any more. Most of all spikes fire between 0.3s ~ 0.7s and this is a good pattern for classification. This may be explained as this: as is mentioned before, after training with PSO, many synapses are weakened strongly while the remainders are strengthened. So many positive feedback loops are weakened and it prevents the *Pathological Synchrony*. Some weights are strengthened greatly and it will make the *Over-Stratification* less possible from happening. However, this experiment is not designed to separate any random spike trains but it is just a demonstration how the PSO enhances the separation property of the LSM and how the PSO prevents the *Pathological Synchrony* and the *Over-Stratification* from happening in the LSM.



8(a) before training



8(b) after training

Fig. 8. The melioration of *Over-Stratification* before and after training with PSO

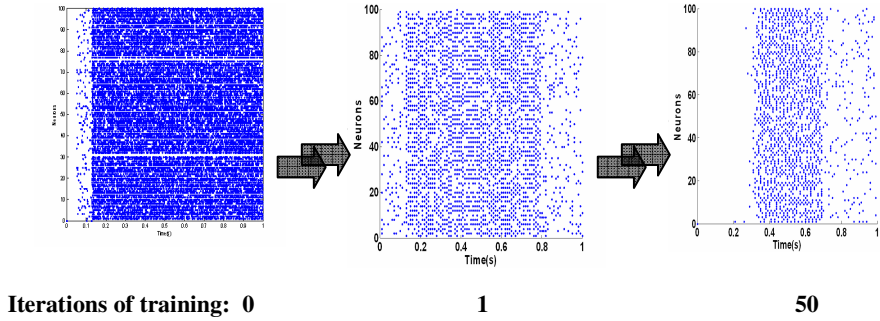


Fig. 9. The Pathological Synchrony is ameliorated after iterative training with PSO. After 50 iterations of training, the neurons no longer fire in no constraint but concentrate in $0.3s \sim 0.7s$. The dots represent the spikes.

4.2 Classification of Fisheriris Data-Set

In this experiment, we test the effect of PSO on the classification of *Fisheriris data-set*. The *Fisheriris data-set* is considered to be a reasonable classification test for the circuit before and after training. The data-set contains 150 cases and each case contains 4 input variables. Before they are fed into the circuit, the variables must be turned to spike trains by Gaussian receptive fields [11]. The 4 input variables converted to 4 spike trains and the trains are fed into the circuit through 4 input neurons. In the readout layer, we adopt the linear regression as the classification algorithm. 90 cases are selected to be training data and the rest to be test data. We also do it with Hebbian-learning as a comparison. The results are shown in Fig. 10. The classification precision is calculated after every 10 iterations. It can be seen that after 70 training iterations, the precision of classification grows gradually from 75% to 95% with PSO but the result of Hebbian-learning doesn't present this trend and it gets a relative worse precision. Simple as the classification algorithm is, but after the enhancement of the circuit, the classification still gets a high precision. This can infer that the separation property of LSM is enhanced greatly.

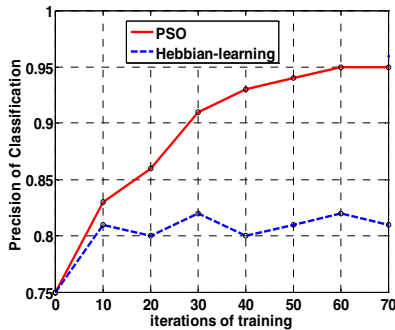


Fig. 10. The result of *Fisheriris data-set* classification as the training process going on

5 Conclusion

The advantage of the Liquid State Machine is that it no longer needs to train the parameters in the recurrent neural network (circuit) but only to train the readout layer, and this makes the application of LSM being quite easy. However, the problems of *Pathological Synchrony* and *Over-Stratification* sometimes result in worse classification precision; our experiment results show that we can do some “optimization” to the randomly-made circuit before we use it. It will help us get more effective classification work.

The random spike trains experiment shows that the separation property of the circuit can be enhanced greatly and the larger separation distance of the liquid state space mean that it will be easier to classify different input streams. And the PSO gets better result than the Hebbian-learning. Also for the problems of *Pathological Synchrony* and *Over-Stratification*, the PSO is shown to be able to overcome them. The *Fisheriris data-set* classification problem is a classical test for the separation property improvement because the separation property affects the classification precision directly. In this case the PSO is useful for the classification problems. Basing on these results, we can conclude that the PSO is effective to enhance the separation property of Liquid State Machine.

The other possible way of further research will be the structure of the circuit because the circuit is created under equation (1). The connection of the circuit is made under a probability. To maximize the separation property, we can use some supervised algorithms or optimization algorithms like ACO or Discrete PSO to decide whether a connection is needed in this circuit.

Acknowledgments. This work was supported in part by the National Nature Science Foundation of China under Grant No. 60674105, High Tech Program of China, No. 2008AA04Z207, the Ph.D. Programs Foundation of Ministry of Education of China, No. 20050487013, and the Nature Science Foundation of Hubei Province of China, No. 2007ABA027.

References

1. Amir, F., Atiya, A.G.P.: New Results on Recurrent Network Training: Unifying the Algorithms and Accelerating Convergence. *IEEE Trans. on Neural Network* 11, 697–709 (2000)
2. Verstraeten, D., Schrauwen, B., et al.: An Experimental Unification of Reservoir Computing Methods. *Neural Network* 2, 391–403 (2007)
3. Jaeger, H.: The "Echo State" Approach to Analysing and Training Recurrent Neural Networks. GMD Report 148, German National Research Center for Information Technology (2001)
4. Maass, W., Natschläger, T., Markram, H.: Real-time Computing Without Stable States: A New Framework for Neural Computation Based on Perturbations. *Neural Computation* 14, 2531–2560 (2002)
5. Steil, J.J.: Backpropagation-decorrelation: Online Recurrent Learning with O(N) Complexity. In: *Proceedings of 2004 IEEE International Joint Conference on Neural Network*, vol. 2, pp. 843–848 (2004)

6. Gerstner, W., Kistler, W.: Spiking Neuron Models. Cambridge University Press, Cambridge (2002)
7. Fang, H.J., Wang, Y.J., Liu, S.: Error-backpropagation with Adaptive Learning Rate in Spiking Neural Network. In: Pre-Proceedings of the Second International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA), pp. 14–17 (2007)
8. Goodman, E.L., Ventura, D.: Spatiotemporal Pattern Recognition via Liquid State Machine. In: Proceedings of the 2006 International Joint Conference on Neural Networks, pp. 3848–3853 (2006)
9. Kennedy, J., Eberhart, R.: Particle Swarm Optimization. In: Proceeding of the 1995 IEEE International Conference on Neural Network, pp. 1942–1948 (2005)
10. Norton, D., Ventura, D.: Preparing More Effective Liquid State Machines Using Hebbian Learning. In: 2006 International Joint Conference On Neural Network, pp. 4243–4248 (2006)
11. Sander, M.B., Han, A.L.P., Joost, N.K.: Error-backpropagation in Temporally Encoded Networks of Spiking Neurons. *Neurocomputing* 48, 17–37 (2002)

A Class of New Large-Update Primal-Dual Interior-Point Algorithms for $P_*(\kappa)$ Linear Complementarity Problems

Huaping Chen, Mingwang Zhang, and Yuqin Zhao

College of Science, China Three Gorges University,
Yichang 443002, China

Abstract. A class of polynomial primal-dual interior-point algorithms for $P_*(\kappa)$ linear complementarity problems (LCPs) are presented. We generalize Ghami et al.'s [A polynomial-time algorithm for linear optimization based on a new class of kernel functions(2008)] algorithm for linear optimization (LO) problem to $P_*(\kappa)$ LCPs. Our analysis is based on a class of finite kernel functions which have the linear and quadratic growth terms. Since $P_*(\kappa)$ LCP is a generalization of LO problem, we lose the orthogonality of the vectors dx and ds . So our analysis is different from the one in Ghami et al.'s algorithm. Despite this, the favorable complexity result is obtained, namely, $O((1 + 2\kappa)n^{\frac{1}{1+p}} \log n \log(n/\epsilon))$, which is better than the usual large-update primal-dual algorithm based on the classical logarithmic barrier function for $P_*(\kappa)$ LCP.

Keywords: Large-update method, Interior-point algorithm, $P_*(\kappa)$ LCPs, Finite kernel function, Polynomial complexity.

1 Introduction

Throughout the paper we consider the LCP as follows:

$$\begin{cases} s = Mx + q, \\ xs = 0, \\ x \geq 0, s \geq 0, \end{cases} \quad (\text{LCP})$$

where $q \in R^n$, $M \in R^{n \times n}$ is a $P_*(\kappa)$ matrix.

LCPs have many applications in mathematical programming and equilibrium problems. The reader can refer to [1]. Since interior-point algorithm was developed, several IPMs of LO have been successfully extended to $P_*(\kappa)$ LCPs [2,3,4,5]. However, it is generally agreed that the large-update IPMs has better practical performance than the small-update IMPs but relatively weak theoretical result. To overcome it many researches haven been done. Recently, M.El Ghami et al. [6] developed a class of kernel functions which take the finite values at the boundary of the feasible region for LO problems, they derived the polynomial complexity for LO with large-update methods.

Inspired by their work, we propose a class of primal-dual IPMs for $P_*(\kappa)$ LCPs based on a class of kernel functions. Of course, it is worth emphasizing that our methods are different from those in [2,3,4]. Our kernel functions are finite, not exponentially convex in the domain and their growth terms are between linear and quadratic. Furthermore, as we lose the orthogonality of the vectors dx and ds , our analysis is also different from that in [6]. Despite this, the favorable complexity result is obtained, namely, $O((1 + 2\kappa)n^{\frac{1}{1+p}} \log n \log(n/\epsilon))$.

This paper is organized as follows. In Section 2 we give some preliminaries. In Section 3 we describe the properties of kernel(barrier) function. In Section 4 we analyze the algorithm and obtain the polynomial complexity of the algorithm. Finally, some concluding remarks are given in Section 5.

We use the following notations throughout the paper: R_+^n denotes the set of n dimensional nonnegative vectors and R_{++}^n , the set of n dimensional positive vectors. For $x = (x_1, x_2, \dots, x_n) \in R^n$, $\|x\|$ is the 2-norm of x , and X is the diagonal matrix from vector x , i.e., $X = \text{diag}(x)$, xs denotes the componentwise product of vectors x and s . e is the n -dimensional vector of ones and I is the n -dimensional identity matrix. J is the index set, i.e., $J = \{1, 2, \dots, n\}$.

2 Preliminaries

2.1 The Properties of $P_*(\kappa)$ Matrix

$P_*(\kappa)$ matrix is introduced by Kojima et al. [7]. Firstly, we give its definitions.

Definition 1. Let κ be a nonnegative number. A matrix $M \in R^{n \times n}$ is called a $P_*(\kappa)$ if

$$(1 + 4\kappa) \sum_{i \in J_+(x)} x_i(Mx)_i + \sum_{i \in J_-(x)} x_i(Mx)_i \geq 0$$

holds for $x \in R^n$, where $J_+(x) = \{i \in J : x_i(Mx)_i \geq 0\}$ and $J_-(x) = \{i \in J : x_i(Mx)_i < 0\}$.

Definition 2. A matrix $M \in R^{n \times n}$ is called a P_* if it is a $P_*(\kappa)$ for some $\kappa \geq 0$, i.e. $P_* = \cup_{\kappa \geq 0} P_*(\kappa)$.

Note that P_* includes the PSD of positive semi-definite matrices, and the class of P-matrices with all the principal minors positive.

2.2 The Central Path and Algorithm

As we all known, the basic idea of primal-dual IPMs is to relax the complementarity condition with the following parameterized system:

$$\begin{cases} s = Mx + q, \\ xs = \mu e, \\ x \geq 0, s \geq 0, \end{cases} \quad (CPP_\mu)$$

where $\mu \geq 0$. Without loss of generality, we assume that (LCP) is strictly feasible. (CPP_μ) has a unique solution for any $\mu \geq 0$. We denote the solution of (CPP_μ) as $(x(\mu), s(\mu))$ for given $\mu > 0$. We also call it a μ -center for given μ and the solution set $\{(x(\mu), s(\mu)) \mid \mu > 0\}$ the central path of the (LCP). As $\mu \rightarrow 0$, the sequence $(x(\mu), s(\mu))$ approaches the solution (x, s) of the (LCP) [7]. We define the following notations:

$$d = \sqrt{\frac{x}{s}}, \quad v = \sqrt{\frac{xs}{\mu}}, \quad dx = \frac{v\Delta x}{x}, \quad ds = \frac{v\Delta s}{s}. \quad (1)$$

Then we have the scaled Newton-system as follows:

$$\begin{cases} -\bar{M}dx + ds = 0, \\ dx + ds = v^{-1} - v, \end{cases} \quad (2)$$

where $\bar{M} = DMD$ and $D = \text{diag}(d)$.

We consider a strictly convex function $\Psi(v)$ which is minimal at $v = e$ and $\Psi(e) = 0$. Then we replace the second equation in (2), by

$$dx + ds = -\nabla\Psi(v). \quad (3)$$

So we get the following modified Newton-system:

$$\begin{cases} -M\Delta x + \Delta s = 0, \\ S\Delta x + X\Delta s = -\mu v\nabla\Psi(v). \end{cases} \quad (4)$$

This system uniquely defines a search direction $(\Delta x, \Delta s)$. The frame of our algorithm is presented as following:

Algorithm 1

Input:

a threshold parameter $\tau > 0$; an accuracy parameter $\epsilon > 0$;
a fixed barrier update parameter θ , $0 < \theta < 1$;
starting point (x^0, s^0) and $\mu^0 > 0$ such that $\Psi(x^0, s^0, \mu^0) \leq \tau$;

begin

$x := x^0$; $s := s^0$; $\mu := \mu^0$;

while $n\mu \geq \epsilon$ do

begin

$\mu := (1 - \tau)\mu$;

while $\Psi(v) > \tau$ do

begin

solve system (4) for Δx and Δs ; determine a step size α ;

update $x := x + \alpha\Delta x$; $s := s + \alpha\Delta s$;

end

end

end

Throughout the paper we assume that a proximity parameter τ and a barrier update parameter θ are given, $\tau = O(n)$ and $0 < \theta < 1$, fixed. We also assume that we are given a strictly feasible point (x, s) which is in a τ -neighborhood of the given μ -center. Then we decrease μ to $\mu_+ = (1 - \theta)\mu$, for some fixed $\theta \in (0, 1)$ and then we solve the modified Newton-system to obtain the unique search direction. The positive condition of a new iterate is ensured by the suitable choice of the step size which is defined by some line search rules. This procedure is repeated until we find a new iterate (x_+, s_+) that is in a τ -neighborhood of the μ_+ -center and then we let $\mu := \mu_+$ and $(x, s) := (x_+, s_+)$. Then μ is again reduced by the factor $1 - \theta$ and we solve the modified Newton-system targeting at the new μ_+ -center, and so on. This process is repeated until μ is small enough, say until $n\mu \leq \varepsilon$. Throughout the paper, we use the proximity function $\Psi(v)$ to find a search direction and to measure the proximity between the current iterates and the μ -center.

3 Properties of the Kernel (Proximity) Function

In this section, we consider a class of univariate functions $\psi(t) : D \rightarrow R_+$, with $R_{++} \subseteq D$ as follows

$$\psi(t) = \frac{t^{p+1} - 1}{p + 1} + \frac{e^{\sigma(1-t)} - 1}{\sigma}, \quad p \in [0, 1], \quad \sigma \geq 1. \tag{5}$$

Obviously, it is different from those in [4-6] for $\lim_{t \downarrow 0} \psi(t) = \psi(0) = \frac{e^\sigma - 1}{\sigma} - \frac{1}{p+1} < \infty$. Specially, for $p = 1$, $\psi(t) = \frac{t^2 - 1}{2} + \frac{e^{\sigma(1-t)} - 1}{\sigma}$ is given in [8].

To simplify the analysis, we define the proximity function as follows

$$\Psi(v) = \sum_{i=1}^n \psi(v_i),$$

which is separable with identical coordinate functions. We call the univariate function $\psi(t)$ the *kernel function* of the proximity function $\Psi(v)$. For $\psi(t)$, the first three derivatives of ψ are given by

$$\psi'(t) = t^p - e^{\sigma(1-t)}, \tag{6}$$

$$\psi''(t) = pt^{p-1} + \sigma e^{\sigma(1-t)}, \tag{7}$$

$$\psi'''(t) = -p(1-p)t^{p-2} - \sigma^2 e^{\sigma(1-t)}. \tag{8}$$

It follows that $\psi(1) = \psi'(1) = 0$, $\psi'(t) > 0 (t > 1)$ and $\psi''(t) > 0 (t > 0)$.

In the following lemma we give several crucial properties which are important in the analysis of the algorithm.

Lemma 1 (Lemma 2.1 in [6]). *Let ψ be as defined in (5), Then,*

$$\psi(\sqrt{t_1 t_2}) \leq \frac{1}{2}(\psi(t_1) + \psi(t_2)), \quad t_1 \geq \frac{1}{\sigma}, \quad t_2 \geq \frac{1}{\sigma}, \tag{9}$$

$$\psi'''(t) < 0. \tag{10}$$

Following [8], we know (9) show that ψ is exponentially convex whenever $t \geq \frac{1}{\sigma}$.

Note that at the start of each outer iteration of the algorithm, just before the update of μ with the factor $1 - \theta$, we have $\Psi(v) \leq \tau$. Due to the update of μ the vector v is divided by the factor $1 - \theta$, with $0 < \theta < 1$, which in general leads to an increase in the value of $\Psi(v)$. Then, during the subsequent inner iterations, $\Psi(v)$ decreases until it passes the threshold τ again. Hence, during the course of the algorithm the largest values of $\Psi(v)$ occur just after the updates of μ . In other words, with $\beta = \frac{1}{\sqrt{1-\theta}}$, we want to find an upper bound for $\Psi(\beta v)$ in terms of $\Psi(v)$. We start with the following lemma.

Lemma 2. *Let $\varrho : [0, \infty) \rightarrow [1, \infty)$ be the inverse function of ψ on $[1, \infty)$. If $\sigma \geq 2$, for all $s > 0$, we have $\varrho(s) \leq 2(s + 1)$.*

Proof. Let $s = \psi(t)$, then $t = \varrho(s)$, assuming $\sigma \geq 2$, one has

$$\frac{t^{1+p} - 1}{1 + p} = \psi(t) + \frac{1 - e^{\sigma(1-t)}}{\sigma} \leq s + \frac{1}{\sigma} \leq s + \frac{1}{2}.$$

Thus

$$t \leq \left((1 + p)\left(s + \frac{1}{2}\right) + 1 \right)^{\frac{1}{1+p}}.$$

Using $0 \leq p \leq 1$, then $\varrho(s) = t \leq (2s + 2)^{\frac{1}{1+p}} \leq 2(s + 1)$. We get the desired result. \square

Theorem 1 (Theorem 3.2 in [9]). *Let ϱ be as defined in Lemma 2. Then for any positive vector v and any $\beta \geq 1$ we have:*

$$\Psi(\beta v) \leq n\psi \left(\beta \varrho \left(\frac{\Psi(v)}{n} \right) \right). \tag{11}$$

Corollary 1. *Let $0 \leq \theta \leq 1$ and $v_+ = \frac{v}{\sqrt{1-\theta}}$. If $\Psi(v) \leq \tau$, then*

$$\Psi(v_+) \leq L := L(n, \theta, \tau) = n\psi \left(\frac{\varrho\left(\frac{\tau}{n}\right)}{\sqrt{1-\theta}} \right) \leq \frac{n}{(1-\theta)^{\frac{p+1}{2}}} \left(\frac{2\tau}{n} + 2 \right)^{p+1}. \tag{12}$$

Proof. It can be easily observed that

$$\psi(t) = \frac{t^{p+1}}{p+1} + \frac{e^{\sigma(1-t)} - 1}{\sigma} \leq \frac{t^{1+p}}{1+p}, \quad \text{for } t \geq 1.$$

Using the result of Lemma 2, by substitution in (11), we obtain

$$L \leq n \left(\frac{\varrho\left(\frac{\tau}{n}\right)}{\sqrt{1-\theta}} \right)^{p+1} \leq n \left(\frac{2\tau}{n} + 2 \right)^{1+p} = \frac{n}{(1-\theta)^{\frac{p+1}{2}}} \left(\frac{2\tau}{n} + 2 \right)^{p+1}.$$

This completes the proof. \square

Remark 1. If $\tau = O(n), \theta = \Theta(1)$, Corollary 1 shows that $L = O(n)$.

Lemma 3 (Lemma 2.7 in [6]). *Suppose that $L \geq 9$, and $\Psi(v) \leq L$. If $\sigma \geq 1 + 2\log(L + 1)$, then $v_i > \frac{3}{2\sigma}$, for all $i = 1, \dots, n$.*

Note that at the start of each inner iteration $\tau < \Psi(v) \leq L$. To ensure that L satisfies the conditions of Lemma 3, from now on we assume that $L \geq 9$, and we choose

$$\sigma = 1 + 2\log(L + 1) \geq 1 + 2\log 10 \approx 5.61. \quad (13)$$

We define the norm-based proximity measure $\delta(v)$ as follows:

$$\delta(v) = \frac{1}{2} \|\nabla \Psi(v)\|. \quad (14)$$

Since $\Psi(v)$ is strictly convex and attains its minimal value zero at $v = e$, we have $\Psi(v) = 0 \Leftrightarrow \delta(v) = 0 \Leftrightarrow v = e$. The following theorem gives a lower bound for $\delta(v)$ in terms of $\Psi(v)$. The reader can refer to the proof in [9].

Theorem 2 (Theorem 4.9 in [9]). *Let ϱ be as defined in Lemma 2. Then*

$$\delta(v) \geq \frac{1}{2} \psi'(\varrho(\Psi(v))).$$

Lemma 4 (Lemma 3.1 in [6]). *If $\Psi(v) \geq 1$, then $\delta(v) \geq \frac{1}{6} \Psi(v)^{\frac{p}{1+p}}$.*

Note that if $\Psi(v) \geq 1$, applying Lemma 4, we can have

$$\delta(v) \geq \frac{1}{6}. \quad (15)$$

4 Analysis of the Algorithm for $P_*(\kappa)$ LCP

4.1 Decrease of the Proximity Function During an Inner Iteration

In this subsection, we will compute a feasible stepsize α and estimate the bound for the decrease of the proximity Function during inner iteration in the form of several lemmas and a theorem. For $P_*(\kappa)$ LCPs, we lose the orthogonality of vector dx and ds , so the analysis is different from that of LO case. After a damped step for fixed μ we have new iterations $x_+ = x + \alpha \Delta x$, $s_+ = s + \alpha \Delta s$. From (11), we have

$$\begin{aligned} x_+ &= x \left(e + \alpha \frac{\Delta x}{x} \right) = x \left(e + \alpha \frac{dx}{v} \right) = \frac{x}{v} (v + \alpha dx), \\ s_+ &= s \left(e + \alpha \frac{\Delta s}{s} \right) = s \left(e + \alpha \frac{ds}{v} \right) = \frac{s}{v} (v + \alpha ds). \end{aligned}$$

Then we get $v_+^2 = x_+ s_+ / \mu = (v + \alpha dx)(v + \alpha ds)$.

Throughout the paper we choose a stepsize α which can ensure the coordinates of the vector $v + \alpha dx$ and $v + \alpha ds$ be positive. We consider the decrease about Ψ as a function of α , denote as

$$f(\alpha) := \Psi(v_+) - \Psi(v) = \Psi(\sqrt{(v + \alpha dx)(v + \alpha ds)}) - \Psi(v).$$

Our aim is to find an upper bound for $f(\alpha)$ by using exponentially convex. In order to do this, we assume for the moment that the step size α is satisfying

$$v_i + \alpha dx_i \geq \frac{1}{\sigma}, \quad v_i + \alpha ds_i \geq \frac{1}{\sigma}, \quad 1 \leq i \leq n. \quad (16)$$

As Ψ is exponentially convex, we have $\Psi(v_+) \leq \frac{1}{2}(\Psi(v + \alpha dx) + \Psi(v + \alpha ds))$. Defining

$$f_1(\alpha) := \frac{1}{2}(\Psi(v + \alpha dx) + \Psi(v + \alpha ds)) - \Psi(v).$$

Obviously, $f(\alpha) \leq f_1(\alpha)$ and one can easily verify that $f(0) = f_1(0) = 0$. The first and second derivative of $f_1(\alpha)$ are the following equations

$$\begin{aligned} f'_1(\alpha) &= \frac{1}{2} \sum_{i=1}^n (\psi'(v_i + \alpha dx_i) dx_i + \psi'(v_i + \alpha ds_i) ds_i), \\ f''_1(\alpha) &= \frac{1}{2} \sum_{i=1}^n (\psi''(v_i + \alpha dx_i) dx_i^2 + \psi''(v_i + \alpha ds_i) ds_i^2). \end{aligned} \quad (17)$$

According to (3) and definition of δ , we have $f'_1(0) = -2\delta(v)^2$.

Since M is a $P_*(\kappa)$ matrix and $M\Delta x = \Delta s$ from (4), for $\Delta x \in R^n$ we have

$$(1 + 4\kappa) \sum_{i \in J_+} \Delta x_i (M\Delta x)_i + \sum_{i \in J_-} \Delta x_i (M\Delta x)_i \geq 0,$$

where $J_+ = \{i \in J : \Delta x_i (M\Delta x)_i \geq 0\}$ and $J_- = J - J_+$. Note that $dx ds = v^2 \Delta x \Delta s / xs = \Delta x \Delta s / \mu$, $\mu > 0$, we can obtain

$$(1 + 4\kappa) \sum_{i \in J_+} dx_i ds_i + \sum_{i \in J_-} dx_i ds_i \geq 0. \quad (18)$$

For notation convenience we define

$$\delta := \delta(v), \quad \sigma_+ = \sum_{i \in J_+} dx_i ds_i, \quad \sigma_- = - \sum_{i \in J_-} dx_i ds_i.$$

Without loss of the generality, we denote

$$v_1 := \min(v).$$

Lemma 5. $\sigma_+ \leq \delta^2$ and $\sigma_- \leq (1 + 4\kappa)\delta^2$.

Proof. By the definition of σ_+ and σ_- , one has

$$\sigma_+ = \sum_{i \in J_+} dx_i ds_i \leq \frac{1}{4} \sum_{i \in J_+} (dx_i + ds_i)^2 \leq \frac{1}{4} \sum_{i=1}^n (dx_i + ds_i)^2 = \frac{1}{4} \|dx + ds\|^2 = \delta^2.$$

Since M is a $P_*(\kappa)$ matrix, from (18), we have

$$(1 + 4\kappa)\sigma_+ - \sigma_- \geq 0.$$

Therefore

$$\sigma_- \leq (1 + 4\kappa)\sigma_+ \leq (1 + 4\kappa)\delta^2,$$

which completes the proof of this lemma. \square

Lemma 6. $\sum_{i=1}^n (dx_i^2 + ds_i^2) \leq 4(1 + 2\kappa)\delta^2$, $\|dx\| \leq 2\sqrt{1 + 2\kappa}\delta$ and $\|ds\| \leq 2\sqrt{1 + 2\kappa}\delta$.

Proof. Since $\delta = \frac{1}{2}\|dx + ds\|$ and $\sum_{i \in J} dx_i ds_i = \sigma_+ - \sigma_-$,

$$2\delta = \|dx + ds\| = \sqrt{\sum_{i=1}^n (dx_i + ds_i)^2} = \sqrt{\sum_{i=1}^n (dx_i^2 + ds_i^2) + 2(\sigma_+ - \sigma_-)}.$$

Using (18), $(1 + 4\kappa)\sigma_+ \geq \sigma_-$, we can have

$$2\delta \geq \sqrt{\sum_{i=1}^n (dx_i^2 + ds_i^2) + 2\left(\frac{1}{1 + 4\kappa}\sigma_- - \sigma_-\right)} = \sqrt{\sum_{i=1}^n (dx_i^2 + ds_i^2) - \frac{8\kappa}{1 + 4\kappa}\sigma_-}.$$

Squaring both sides, we have

$$4\delta^2 + \frac{8\kappa}{1 + 4\kappa}\sigma_- \geq \sum_{i=1}^n (dx_i^2 + ds_i^2).$$

By Lemma 7, we can obtain

$$4(1 + 2\kappa)\delta^2 \geq 4\delta^2 + \frac{8\kappa}{1 + 4\kappa}\sigma_- \geq \sum_{i=1}^n (dx_i^2 + ds_i^2).$$

So

$$2\sqrt{1 + 2\kappa}\delta \geq \sqrt{\sum_{i=1}^n (dx_i^2 + ds_i^2)} \geq \|dx\|$$

holds. And we can get $2\sqrt{1 + 2\kappa}\delta \geq \|ds\|$ in the same way. This completes the proof. \square

Lemma 7. $f_1''(\alpha) \leq 2(1 + 2\kappa)\delta^2\psi''(v_1 - 2\alpha\sqrt{1 + 2\kappa}\delta)$.

Proof. The proof is simple, the reader can refer to [2]. \square

Lemma 8. $f_1'(\alpha) \leq 0$ if α is satisfying

$$-\psi'(v_1 - 2\alpha\delta\sqrt{1 + 2\kappa}) + \psi'(v_1) \leq \frac{2\delta}{\sqrt{1 + 2\kappa}}. \quad (19)$$

Proof. Applying $f_1'(0) = -2\delta(v)^2$ and Lemma 7, one can easily prove this lemma. \square

Lemma 9 (Lemma 4.5 in [2]). Let $\rho : [0, \infty) \rightarrow (0, 1]$ denotes the inverse function of the restriction of $-\frac{1}{2}\psi'(t)$ to the interval $(0, 1]$. Then the largest step size α that satisfying (19) is given by

$$\bar{\alpha} := \frac{1}{2\delta\sqrt{1+2\kappa}} \left(\rho(\delta) - \rho \left(\left(1 + \frac{1}{\sqrt{1+2\kappa}} \right) \delta \right) \right). \quad (20)$$

Lemma 10 (Lemma 4.6 in [3]). Let ρ and $\bar{\alpha}$ as defined in Lemma 9. Then for $a = 1 + \frac{1}{\sqrt{1+2\kappa}}$, we have

$$\bar{\alpha} \geq \frac{1}{1+2\kappa} \frac{1}{\psi''(\rho(a\delta))}. \quad (21)$$

Defining

$$\tilde{\alpha} = \frac{1}{1+2\kappa} \frac{1}{\psi''(\rho(a\delta))}, \quad (22)$$

then according to Lemma 10, we can have $\bar{\alpha} \geq \tilde{\alpha}$.

Lemma 11 (Lemma 4.8 in [2]). If the stepsize α is subject to $\alpha \leq \bar{\alpha}$, then $f(\alpha) \leq -\alpha\delta^2$.

Theorem 3. Let ρ as defined in Lemma 9 and $\tilde{\alpha}$ as stated in (22) and $\Psi(v) \geq 1$. Then

$$f(\tilde{\alpha}) \leq -\frac{1}{1+2\kappa} \frac{\delta^2}{\psi''(\rho(a\delta))} \leq -\frac{\delta}{2(1+2\kappa)(6+a)\sigma}. \quad (23)$$

Proof. Since $\tilde{\alpha} \leq \bar{\alpha}$, by Lemma 11, we have $f(\tilde{\alpha}) \leq -\tilde{\alpha}\delta^2$, where $\tilde{\alpha} = \frac{1}{1+2\kappa} \frac{1}{\psi''(\rho(a\delta))}$. Thus the first inequality follows. According to the definition of ρ , we put $t = \rho(a\delta)$ for $\frac{1}{\sigma} < t < 1$. This implies $-\frac{1}{2}\psi'(t) = a\delta$. Using (6) and $t \leq 1$, we get

$$e^{\sigma(1-t)} = 2a\delta + t^p \leq 2a\delta + 1, \quad \text{for } p \in [0, 1]. \quad (24)$$

And using $t \geq \frac{1}{\sigma}$, as well as $p \in [0, 1]$, we obtain

$$\begin{aligned} \tilde{\alpha} &= \frac{1}{1+2\kappa} \frac{1}{\psi''(t)} = \frac{1}{1+2\kappa} \frac{1}{pt^{p-1} + \sigma e^{\sigma(1-t)}} \\ &\geq \frac{1}{1+2\kappa} \frac{1}{p\sigma^{1-p} + \sigma e^{\sigma(1-t)}} \geq \frac{1}{1+2\kappa} \frac{1}{\sigma(1 + e^{\sigma(1-t)})}. \end{aligned}$$

Furthermore, using (15) and (24), we have

$$\begin{aligned} \tilde{\alpha} &\geq \frac{1}{1+2\kappa} \frac{1}{\sigma(2+2a\delta)} = \frac{1}{1+2\kappa} \frac{1}{2\sigma(1+a\delta)} \\ &\geq \frac{1}{1+2\kappa} \frac{1}{2\sigma(6\delta+a\delta)} = \frac{1}{2(1+2\kappa)(6+a)\sigma\delta}. \end{aligned}$$

Hence

$$f(\tilde{\alpha}) \leq -\frac{\delta^2}{2(1+2\kappa)(6+a)\sigma\delta} = -\frac{\delta}{2(1+2\kappa)(6+a)\sigma}.$$

The proof is completed. \square

In what follows we use notation

$$\hat{\alpha} = \frac{1}{2(1+2\kappa)(6+a)\sigma\delta}.$$

and we also use $\hat{\alpha}$ as our default step size. In fact, $\hat{\alpha}$ satisfies (16). Using Lemma 3, Lemma 6, and the definition of v_1 , we obtain

$$\begin{aligned} v_i + \hat{\alpha}dx_i &\geq v_1 - 2\hat{\alpha}\sqrt{1+2\kappa}\delta \geq \frac{3}{2\sigma} - \frac{2\sqrt{1+2\kappa}\delta}{2(1+2\kappa)(6+a)\sigma\delta} \\ &= \frac{3}{2\sigma} - \frac{1}{(6+a)\sigma\sqrt{1+2\kappa}} > \frac{1}{\sigma}. \end{aligned}$$

So does $v_i + \hat{\alpha}ds_i$.

According to Lemma 4 and (23), the following inequality holds:

$$f(\tilde{\alpha}) \leq -\frac{\delta}{2(1+2\kappa)(6+a)\sigma} \leq -\frac{\Psi_0^{\frac{p}{1+p}}}{12(1+2\kappa)(6+a)\sigma}. \quad (25)$$

4.2 Iteration Bound of Large-Update Method

In this subsection we analyze the complexity of the algorithm. We cite the following lemma in [10] to obtain iteration bound for the algorithm.

Lemma 12. *Let t_0, t_1, \dots, t_K be a sequence of positive numbers such that $t_{k+1} \leq t_k - \beta t_k^{1-\gamma}$, $k = 0, \dots, K-1$, where $\beta > 0$ and $0 < \gamma \leq 1$. Then $K \leq \lceil t_0^\gamma / \beta \gamma \rceil$.*

Lemma 13. *If K denotes the number of inner iteration, then we have*

$$K \leq \frac{192(1+2\kappa)\sigma n^{\frac{1}{1+p}}(\frac{2\tau}{n} + 2)}{\sqrt{1-\theta}}.$$

Proof. By (25), we have $\Psi_{k+1} \leq \Psi_k - \eta\Psi_k^{1-\gamma}$, $k = 0, 1, 2, \dots, K-1$, with $\eta = \frac{1}{12(1+2\kappa)(6+a)\sigma}$, $\gamma = \frac{1}{1+p}$. Applying Lemma 12 and $a < 2$ yields $K \leq 96(1+2\kappa)(1+p)\sigma\Psi_0^{\frac{1}{1+p}}$. Using Corollary 1 and $p \leq 1$, one can easily obtain the result. \square

The number of outer iterations is bounded above by $\frac{1}{\theta} \log \frac{n}{\epsilon}$, seeing [9]. By multiplying the number of outer iterations and the number of inner iterations, we can get an upper bound for the total number of iterations, namely

$$\frac{192(1+2\kappa)\sigma n^{\frac{1}{p+1}}}{\theta\sqrt{1-\theta}} \left(\frac{2\tau}{n} + 2 \right) \log \frac{n}{\epsilon}.$$

Theorem 4. *Let $\tau = O(n)$, $\theta = \Theta(1)$, $\sigma = O(\log n)$, which are characteristics of the large-update methods, the **Algorithm 1** will obtain ϵ -approximate solutions of (LCP) after at most $O\left((1+2\kappa)n^{\frac{1}{p+1}} \log n \log \frac{n}{\epsilon}\right)$ iterations.*

Remark 2. For $p = 1$, and $\kappa = 0$, the iteration bound is $O(\sqrt{n} \log n \log \frac{n}{\epsilon})$, which is the currently best iteration bound for LO problem with the large-update methods. For $p = 1, \kappa > 0$, the iteration bound is $O((1 + 2\kappa)\sqrt{n} \log n \log \frac{n}{\epsilon})$, which is the currently best iteration bound for $P_*(\kappa)$ LCPs based on the self-regular function.

5 Concluding Remarks

In this paper, we generalize Ghami et al.'s [6] algorithm for linear optimization(LO) problem to $P_*(\kappa)$ LCPs and give the **Algorithm 1** for large- update method based a class of finite kernel functions. In section 4, we develop favorable complexity result for our algorithm. The numerical results need further study. Also the extensions to Nonlinear complementarity problems are to be investigated.

Acknowledgment. Supported by Natural Science Foundation of Educational Commission of Hubei Province of China (NO. D200613009).

References

1. Bullups, S.C., Murty, K.G.: Complementarity problems. J. Comput. Appl. Math. 124, 303–318 (2000)
2. Cho, G.M.: A new large-update interior point algorithm for $P_*(\kappa)$ linear complementarity problems. J. Comput. Appl. Math. 216, 265–278 (2008)
3. Cho, G.M., Kim, M.K.: A new large-update interior-point algorithm for $P_*(\kappa)$ LCPs based on kernel function. Appl. Math. Comput. 182, 1169–1183 (2006)
4. Cho, G.M., Kim, M.K., Lee, Y.H.: Complexity of large-update interior point algorithm for $P_*(\kappa)$ linear complementarity problem. Comput. Math. Appl. 53, 948–960 (2007)
5. Illés, T., Nagy, M.: The Mizuno - Todd - Ye predictor - corrector algorithm for sufficient matrix linear complementarity problem. European J. Oper. Res. 181, 1097–1111 (2007)
6. El Ghami, M., Ivanov, I., Melissen, J.B.M., Roos, C., Steihaug, T.: A polynomial-time algorithm for linear optimization based on a new class of kernel functions. J. Comput. Appl. math. (2008), <http://www.sciencedirect.com>
7. Kojima, M., Megiddo, N., Noma, T., Yoshise, A.: A Unified Approach to Interior Point Algorithms for Linear Complementarity Problems. LNCS, vol. 538. Springer, Heidelberg (1991)
8. A new efficient large-update primal-dual interior-point method based on a finite barrier. SIAM. J. Optim. 13, 766–782 (2003)
9. Bai, Y.Q., El Ghami, M., Roos, C.: A comparative study of kernel functions for primal-dual interior-point algorithms in linear optimization. SIAM. J. Optim. 15, 101–128 (2004)
10. Peng, J., Roos, C., Terlaky, T.: Self-Regularity: A New Paradigm for Primal-Dual Interior-Point Algorithms. Princeton University Press, USA (2002)

A Novel Artificial Immune System for Multiobjective Optimization Problems

Jiaquan Gao and Lei Fang

Zhijiang College, Zhejiang University of Technology, Hangzhou 310024, China
gaojiaquan@gmail.com

Abstract. This study presents a novel weight-based multiobjective artificial immune system (WBMOAIS) based on opt-aiNET. The proposed algorithm follows the elementary structure of opt-aiNET, but has the following distinct characteristics: At first, a randomly weighted sum of multiple objectives is used as a fitness function; Secondly, the individuals of the population are chosen from the memory, which is a set of elite solutions. Lastly, in addition to the clonal suppression algorithm similar to that used in opt-aiNET, a new truncation algorithm with similar individuals (TASI) is presented in order to eliminate the similar individuals in memory and obtain a well-distributed spread of non-dominated solutions. Simulation results show WBMOAIS outperforms the vector immune algorithm (VIS) and the elitist non-dominated sorting genetic system (NSGA-II).

Keywords: Multiobjective optimization, Artificial immune system, Similar individuals, Evolutionary algorithm.

1 Introduction

For the multiobjective optimization problems (MOOPs), evolutionary algorithms (EAs) in general have been demonstrated to be effective and efficient tools for finding approximations of the Pareto front. For a good overview of the current state-of-the-art in multiobjective evolutionary algorithms (MOEAs), we refer the reader to some of the main books in the field [1,2].

During the last decade, based on principles of the immune system, a new paradigm, called artificial immune system (AIS), has been employed for developing interesting algorithms in many fields such as pattern recognition, computer defense, optimization etc. [3,4]. However, very few direct approaches to MOOPs using AIS have been proposed, and most of the existing work considers the use of AIS as a tool for keeping diversity in the population of a genetic algorithm (GA) [5] or handling constraints in EAs [6]. The first reported approach which uses AIS for solving MOOPs was proposed by Yoo and Hajela (1999) [7]. In their approach, AIS is used for modifying the fitness values of a GA. Although Yoo and Hajela's algorithm cannot be considered a true multiobjective artificial immune system (MOAIS), it is a pioneer in using AIS ideas in MOOPs. Coello Coello and Cruz Cortés in 2002 presented a MOAIS based on the clonal

selection theory [8]. The algorithm, called the multiobjective immune system algorithm (MISA), can be considered the really first attempt to solve MOOPs directly with AIS. The performance of MISA has been improved in further work of the same authors in 2005 [9]. In the following year, based on opt-aiNET, the multi-modal AIS optimization algorithm proposed by Castro and Timmis [10], Freschi and Repetto presented a vector immune system (VIS) [11]. In the Freschi and Repetto's study, VIS follows the elementary structure of the opt-aiNET optimization algorithm, and the differences between opt-aiNET and VIS are very few. Besides them, many approaches using the AIS metaphor have been presented in recent years. The representatives of them include Luh and Chueh's multiobjective immune algorithm (MOCSA) [12], the immune dominance clonal multiobjective algorithm (IDCMA) presented by Jiao et al. [13], the immune forgetting multiobjective optimization algorithm (IFMOA) suggested by Wang et al. [14], the adaptive clonal selection algorithm for multiobjective optimization (ACSAMO) proposed by Wang and Mahfouf [15], and Zhang's multiobjective optimization immune algorithm in dynamic environments [16].

In this study, like VIS, we follow the elementary structure of opt-aiNET and present a novel multiobjective artificial immune algorithm. Compared to the other MOAIS based on opt-aiNET, our proposed algorithm, called the weight-based multiobjective artificial immune system (WBMOAIS), has its distinct features. Firstly, WBMOAIS uses a random weighted sum of multiple objectives as a fitness function. Secondly, we define a term called similar individuals. Based on the definition, a new diversity approach, named truncation algorithm with similar individuals (TASI) is presented. Here we use two diversity approaches together, TASI and the clonal suppression algorithm which is similar to what used in opt-aiNET, to eliminate similar cells in memory. Furthermore, TASI is used as the main diversity approach in order to obtain a better distribution of Pareto-optimal solutions. In addition, the individuals of the population are chosen from the memory. Our proposed algorithm, WBMOAIS, is tested on three standard problems and compared with VIS and NSGA-II [17].

2 Multiobjective Optimization Problem

A MOOP has a number of objective functions which are to be minimized or maximized. Without loss of generality, here the minimization for each objective is considered. In the following, we state the MOOP in its general form:

$$\text{Minimize } f_i(\mathbf{x}), \quad i = 1, 2, \dots, m \quad (1)$$

where each $f_i, 1 \leq i \leq m$, is an objective function, $\mathbf{x} = [x_1, x_2, \dots, x_n]^T \in \Omega$ is the vector of decision variables and $\Omega \subseteq \mathcal{R}^n$ is the domain of the variables, defined by their lower and upper bounds: $x_i^L \leq x_i \leq x_i^U, i = 1, 2, \dots, n$. The feasible set $\mathcal{F} \subset \Omega$ can be restricted by inequality and equality constraints: $g_j(\mathbf{x}) \geq 0, j = 1, 2, \dots, J$ and $h_k(\mathbf{x}) = 0, k = 1, 2, \dots, K$.

Next, we assume that the vector $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x})]^T$ and give several definitions related to MOOPs.

Definition 1. (Pareto Dominance): A feasible decision vector \mathbf{x}_p is said to dominate another feasible vector \mathbf{x}_q (denoted by $\mathbf{x}_p \prec \mathbf{x}_q$), if both conditions (i) and (ii) are true.

- (i) \mathbf{x}_p is no worse than \mathbf{x}_q in all objectives: $\forall i = 1, 2, \dots, m \quad f_i(\mathbf{x}_p) \leq f_i(\mathbf{x}_q)$.
- (ii) \mathbf{x}_p is strictly better than \mathbf{x}_q in at least one objective: $\exists i = 1, 2, \dots, m \quad f_i(\mathbf{x}_p) < f_i(\mathbf{x}_q)$.

If there is no solution \mathbf{x}_p that dominates \mathbf{x}_q , then \mathbf{x}_q is a Pareto optimal solution.

Definition 2. (Pareto Optimal Set): For a given MOOP, the Pareto optimal set, \mathcal{PS} , is defined as $\mathcal{PS} := \{\mathbf{x} \in \mathcal{F} \mid \neg \exists \mathbf{x}^* \in \mathcal{F}, \mathbf{x}^* \prec \mathbf{x}\}$.

Definition 3. (Pareto Front): For a given MOOP and the Pareto optimal set, the Pareto front, \mathcal{PF} , is defined as $\mathcal{PF} := \{\mathbf{f}(\mathbf{x}) \mid \mathbf{x} \in \mathcal{PS}\}$.

In general, a Pareto-optimal set is always a non-dominated set. But the non-dominated solutions found by an optimization algorithm are not certain to be able to represent the true Pareto-optimal set. Therefore, In this study the actual Pareto front, termed $\mathcal{PF}_{\text{true}}$, is distinguished from the final set of non-dominated solutions returned by a MOEA, termed $\mathcal{PF}_{\text{known}}$.

3 The Proposed Algorithm

In this section, we will present a novel multiobjective optimization algorithm. The main procedures are listed as follows:

- (1) A random initial population P_0 of size N_{pop} is created. Set the memory $Q_0 = \emptyset$ and its maximum size = N_{mem} . Set a counter $t = 1$, and set a flag $N_{\text{flag}} = 0$ (one of conditions to decide to go to (8) or (9)).
- (2) For each cell in the population, do
 - (a) Randomly specify the weight values w_1, w_2, \dots, w_m , where $w_1 + w_2 + \dots + w_m = 1$, and $w_i \in [0, 1], i = 1, 2, \dots, m$.
 - (b) Reproduce the cell N_{clones} copies and mutate each clone by a random perturbation (see Sect. 3.1).
- (3) Compute the non-dominated individuals among offspring and memory cells, and copy them to the memory Q_{t+1} .
- (4) If the size of Q_{t+1} exceeds $N_{\text{threshold}}$ (a threshold of the memory size), then set $N_{\text{flag}} = 1$ and continue. Otherwise, go to (6).
- (5) Apply the clonal suppression algorithm (see Sect. 3.3) in Q_{t+1} to eliminate those memory clones whose affinity with each other is less than a pre-specified threshold.
- (6) If $|Q_{t+1}| > N_{\text{mem}}$, then perform TASI (see Sect. 3.2).
- (7) If t can be exactly divided by N_{in} (inner loop times) and $N_{\text{flag}} = 1$, then set $N_{\text{flag}} = 0$ and go to (9). Otherwise, continue.
- (8) If $|Q_{t+1}| \geq N_{\text{pop}}$, then randomly select N_{pop} cells from Q_{t+1} to create the population P_{t+1} . Otherwise, copy the cells of Q_{t+1} to P_{t+1} and then randomly select $N_{\text{pop}} - |Q_{t+1}|$ cells from the dominated individuals among offspring and memory cells and add them to P_{t+1} . Go to (10).

- (9) If the size of Q_{t+1} exceeds $(1 - p_{\text{rand}})N_{\text{pop}}$, then randomly select $(1 - p_{\text{rand}})N_{\text{pop}}$ cells from Q_{t+1} to create the population P_{t+1} , and then randomly generate $p_{\text{rand}}N_{\text{pop}}$ new cells and add them to the population P_{t+1} . Otherwise, the cells in Q_{t+1} are copied to P_{t+1} and new randomly generated cells to fill the remaining population.
- (10) If $t > T$ (the maximum reproduction generation) or other termination condition is satisfied, then terminate the algorithm and output the non-dominated individuals in the memory. Otherwise, $t = t + 1$ and return to (2).

3.1 Mutation Operator

The mutation is performed according to the following expression: $\mathbf{x}' = \mathbf{x} + \alpha N(0, 1)$, where $\alpha = \beta \exp(-f^*)$. \mathbf{x} is the real-valued vector, \mathbf{x}' is its mutation version. $N(0, 1)$ is a vector of Gaussian random numbers of mean 0 and standard deviation 1. f^* is the normalized value of f (where f is a weighted sum of objectives), determined by $f^* = (f - f_{\min}) / (f_{\max} - f_{\min})$. The value β is a parameter of the algorithm and it is chosen to define the maximum amplitude of mutation. If the real-valued vectors \mathbf{x}' and \mathbf{x} are defined in a normalized parameter space, β is within the range $[0, 1]$.

3.2 Truncation Algorithm with Similar Individuals (TASI)

Here, we present TASI, whose aim is to reduce the size of the memory Q_{t+1} of size N'_{mem} to N_{mem} (where $N'_{\text{mem}} > N_{\text{mem}}$). First, let us give the definition of similar individuals before proposing this algorithm.

Definition 4. (Similar Individuals): For a given MOOP and the solution set P , the solutions P_1 ($P_1 \in P$) and P_2 ($P_2 \in P$) are referred to as similar individuals, if their Euclidean distance in the objective space is the shortest in set P .

Given the above definition of similar individuals, our new truncation algorithm, TASI, is listed as follows.

- (1) Calculate the Euclidean distance (in objective space) between any two solutions i and j in the memory Q_{t+1} .
- (2) Select two similar individuals i and j from Q_{t+1} and remove one of them from Q_{t+1} . The solution i is chosen for removal if the following condition is true. $\exists 0 < k < |Q_{t+1}|$, such that $d_i^k < d_j^k$, and $\forall 0 < l < k, d_i^l = d_j^l$, where d_i^k denotes the distance of i to its k -th nearest neighbor in Q_{t+1} . Otherwise, the solution j is chosen for removal.
- (3) If $|Q_{t+1}| > N_{\text{mem}}$, then return to (2). Otherwise, terminate the procedure.

3.3 Clonal Suppression Algorithm

The clonal suppression algorithm is to eliminate similar cells in the memory [3]. The detailed procedure is described as follows:

- (1) *Computation of affinity*: Calculate the affinity between two cells v and w according to $ay(v, w) = \sqrt{\sum_{i=1}^m (f_i^v - f_i^w)^2}$, where f_i^v denotes the i -th objective value of the cell v . It can be seen that $ay(v, w)$ is the Euclidean distance (in the objective space) between cells v and w .
- (2) *Computation of concentration*: For any antibody v , its concentration number is calculated by the following equation: $ce(v) = \frac{1}{|Q_{t+1}|} \cdot \sum_{w=1}^{|Q_{t+1}|} ac(v, w)$, where if $ay(v, w) < \delta_1$, $ac(v, w) = 1$. Otherwise, $ac(v, w) = 0$. δ_1 ($0 < \delta_1 \leq 1$) is a given threshold.
- (3) *Suppression of cells*: By step (2), we can obtain the concentration number of each cell in the memory Q_{t+1} , and then eliminate the cells whose concentration number is more than the given threshold δ_2 ($0 < \delta_2 \leq 1$).

For WBMOAIS, the task of the algorithm is to eliminate the most similar cells in the memory, so the threshold δ_1 usually is a smaller value (e.g., $\delta_1 = 10^{-7}$).

4 Experiments

In this section, the proposed algorithm WBMOAIS is compared against two state-of-the-art algorithms: NSGA-II and VIS. All experiments are conducted on an IBM computer, which is equipped with a Pentium IV 2.8G processor and 1 GB of internal memory. The operating system is Windows 2000 server and the programming language is C++. The compiler is Borland C++ 6.0.

4.1 Experimental Setting

Here we list the specification of parameters for all three algorithms. For WBMOAIS, the results indicated below were obtained using the following parameters: population size $N_{\text{pop}} = 100$, size of external memory $N_{\text{mem}} = 200$, number of clones for each cell $N_{\text{clones}} = 5$, number of inner iterations $N_{\text{in}} = 5$, threshold of external memory size $N_{\text{threshold}} = 400$, percentage of random cells at each outer iteration $p_{\text{rand}} = 20\%$, and $\beta = 0.85$.

For VIS, the same parameters as in its original literature are used except that the size of external memory is 200. The population size is 100, the number of inner iterations is 5, the number of clones for each cell is 5, the percentage of random cells at each outer iteration is 20%, and σ_{start} is 0.1.

For NSGA-II, we maintain the same parameters reported in its original literature, which include a population size of 100, a crossover rate of 0.9, and a mutation rate of $1/N_{\text{vars}}$. where N_{vars} = number of decision variables.

4.2 Performance Measures

Here two different measures (Spacing (**S**), and Generational Distance (**GD**)) are chosen and are used for numerical comparison of the non-dominated fronts produced by the algorithms; each of them takes into account a particular desired characteristic of $\mathcal{PF}_{\text{known}}$.

- (1) **Spacing (S):** To measure how well the solutions throughout $\mathcal{PF}_{\text{known}}$ are distributed, we adopt the metric suggested by Schott (1995) [11]. The metric is calculated by $S = \sqrt{\frac{1}{|\mathcal{PF}_{\text{known}}|} \sum_{i=1}^{|\mathcal{PF}_{\text{known}}|} (d_i - \bar{d})^2}$, where $\bar{d} = \sum_{i=1}^{|\mathcal{PF}_{\text{known}}|} d_i / |\mathcal{PF}_{\text{known}}|$ and $d_i = \min_{k \in \mathcal{PF}_{\text{known}} \wedge k \neq i} \sum_{j=1}^m |f_j^i - f_j^k|$. The distance measure is the minimum value of the sum of the absolute difference in objective function value between the i -th solution and any other solution in the obtained non-dominated set.
- (2) **Generational Distance (GD):** In order to measure the convergence of $\mathcal{PF}_{\text{known}}$, the metric suggested by Veldhuizen (1999) [11] is used. Assume that \mathcal{P}^* is a known Pareto-optimal set. Instead of finding whether a solution of $\mathcal{PF}_{\text{known}}$ belongs to the set \mathcal{P}^* or not, this metric finds an average distance of the solutions of $\mathcal{PF}_{\text{known}}$ from \mathcal{P}^* , as follows: $GD = (\sum_{i=1}^{|\mathcal{PF}_{\text{known}}|} d_i^p)^{1/p} / |\mathcal{PF}_{\text{known}}|$. For $p = 2$, the parameter d_i is the Euclidean distance (in the objective space) between the solution $i \in \mathcal{PF}_{\text{known}}$ and the nearest member of \mathcal{P}^* : $d_i = \min_{k=1, \dots, |\mathcal{P}^*|} \sqrt{\sum_{j=1}^m [f_j^{(i)} - f_j^{*(k)}]^2}$, where $(f_j^*)^k$ is the j -th objective function value of the k -th member of \mathcal{P}^* .

4.3 Test Problems

In order to validate our approach, we choose three test problems from a number of significant past studies in this area. They are ZDT6 suggested by Zitzler et al., POL proposed by Poloni et al., and Viennet’s VNT [11].

Test problem 1: ZDT6. The first test problem was proposed by Zitzler et al. (2000) [11]:

$$\text{ZDT6: min} \begin{cases} f_1(\mathbf{x}) &= 1 - \exp(-4x_1) \sin^6(6\pi x_1), \\ f_2(\mathbf{x}) &= g(\mathbf{x})[1 - (f_1(\mathbf{x})/g(\mathbf{x}))^2]. \\ g(\mathbf{x}) &= 1 + 9[(\sum_{i=2}^{10} x_i)/9]^{0.25} \end{cases} \quad (2)$$

where $x_i \in [0, 1], i = 1, 2, \dots, 10$. The Pareto-optimal region corresponds to $x_1^* \in [0, 1]$ and $x_i^* = 0$ for $i = 2, 3, \dots, 10$. For this test problem, the adverse density of solutions across the Pareto-optimal front, coupled with the nonconvex nature of the front, may cause difficulties for many multiobjective optimization algorithms to converge to the true Pareto-optimal front.

In this case, all three algorithms stop after 40000 fitness function evaluations. The comparison of results between the true Pareto front of ZDT6 and the Pareto front produced by WBMOAIS, VIS and NSGA-II is shown in Fig. 11. The values of the two metrics for each algorithm are presented in Table 11.

From Fig. 11, we can observe that WBMOAIS and VIS are able to approximate the true Pareto front but WBMOAIS shows better spread of solutions than VIS. For NSGA-II, it is seen that NSGA-II has a very uniform spread of solutions but it shows difficulties in detecting the global Pareto front, getting stuck at a local one. These are confirmed by the results in Table 11. From Table 11, it is observed that WBMOAIS shows better behavior for all metrics than VIS and NSGA-II except that NSGA-II has a small standard deviation of the S metric.

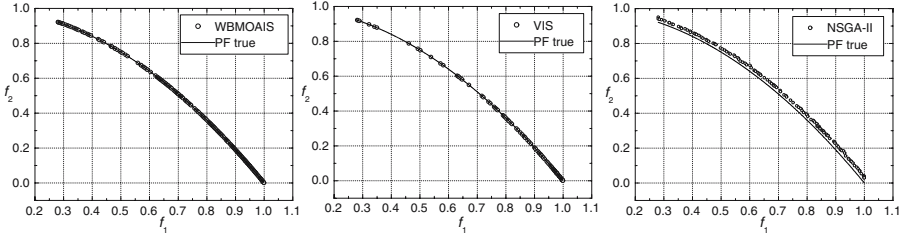


Fig. 1. Pareto-optimal front obtained using WBMOAIS, VIS and NSGA-II for ZDT6

Table 1. Results for ZDT6: mean value and standard deviation (σ)

	WBMOAIS	VIS	NSGA-II
$S(\text{mean})$	<u>4.29E-03</u>	9.84E-03	5.37E-03
$S(\sigma)$	<u>5.71E-04</u>	2.76E-03	<u>3.86E-04</u>
$GD(\text{mean})$	<u>1.32E-05</u>	1.82E-05	1.28E-03
$GD(\sigma)$	<u>2.93E-07</u>	2.59E-06	1.81E-04

Test problem 2: POL. The test problem was proposed by Poloni et al. (2000) [1] and has been used by many researchers subsequently:

$$\text{POL: min} \begin{cases} f_1(\mathbf{x}) &= 1 + (A_1 - B_1)^2 + (A_2 - B_2)^2, \\ f_2(\mathbf{x}) &= (x_1 + 3)^2 + (x_2 + 1)^2, \\ A_1 &= 0.5 \sin 1 - 2 \cos 1 + \sin 2 - 1.5 \cos 2, \\ A_2 &= 1.5 \sin 1 - \cos 1 + 2 \sin 2 - 0.5 \cos 2, \\ B_1 &= 0.5 \sin x_1 - 2 \cos x_1 + \sin x_2 - 1.5 \cos x_2, \\ B_2 &= 1.5 \sin x_1 - \cos x_1 + 2 \sin x_2 - 0.5 \cos x_2. \end{cases} \quad (3)$$

where $-\pi \leq x_1, x_2 \leq \pi$. The function has a nonconvex and disconnected Pareto-optimal set. Like other problems having disconnected Pareto-optimal sets, **POL** may also cause difficulty to many multiobjective optimization algorithms.

The total number of fitness function evaluations for all three algorithms has been set to 3000 in this case. The comparison of results between the true Pareto front of POL and the Pareto front produced by WBMOAIS, VIS and NSGA-II is shown in Fig 2. The values of the two metrics for each algorithm are presented in Table 2.

From Fig 2, it can be observed that all three algorithms are able to approximate the true Pareto front. However, compared to VIS and NSGA-II, WBMOAIS has a better spread of solutions. These can be also obtained from the results in Table 2. From Table 2, we can find that WBMOAIS has a smallest value for all metrics, and thus outperforms VIS and NSGA-II with respect to the two metrics.

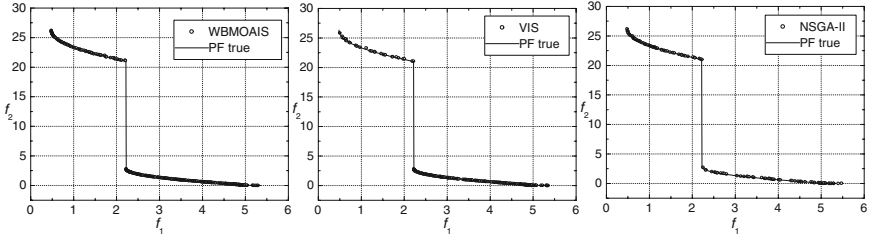


Fig. 2. Pareto-optimal front obtained using WBMOAIS, VIS and NSGA-II for POL

Table 2. Results for POL: mean value and standard deviation (σ)

	WBMOAIS	VIS	NSGA-II
$S(\text{mean})$	<u>3.47E-02</u>	4.62E-02	6.23E-02
$S(\sigma)$	<u>5.24E-03</u>	7.94E-03	1.14E-02
$GD(\text{mean})$	<u>8.26E-04</u>	1.21E-03	3.20E-03
$GD(\sigma)$	<u>1.69E-04</u>	1.79E-04	5.75E-04

Test problem 3: VNT. The test problem was presented by Viennet (1996) [11]:

$$\text{VNT: min} \begin{cases} f_1(\mathbf{x}) &= 0.5(x_1^2 + x_2^2) + \sin(x_1^2 + x_2^2), \\ f_2(\mathbf{x}) &= (3x_1 - 2x_2 + 4)^2/8 + (x_1 - x_2 + 1)^2/27 + 15, \\ f_3(\mathbf{x}) &= (x_1^2 + x_2^2 + 1)^{-1} - 1.1 \exp(-x_1^2 - x_2^2). \end{cases} \quad (4)$$

where $-3 \leq x_1, x_2 \leq 3$. This problem has two variables, and presents several challenging features, such as a high dimensional objective space, discontinuous Pareto optimal set and several local minima in the objective functions.

For VNT, algorithms stop after 4000 fitness function evaluations. The comparison of results between the true Pareto front of VNT and the Pareto front produced by WBMOAIS, VIS and NSGA-II is shown in Fig 3. Table 3 presents the values of the two metrics for each algorithm.

In this case, from Table 3, it can be found that WBMOAIS has the smallest value for all metrics and show better performance than VIS and NSGA-II. These can also be observed by Fig 3.

Table 3. Results for VNT: mean value and standard deviation (σ)

	WBMOAIS	VIS	NSGA-II
$S(\text{mean})$	<u>2.37E-02</u>	5.15E-02	4.76E-02
$S(\sigma)$	<u>8.78E-03</u>	9.82E-03	9.92E-03
$GD(\text{mean})$	<u>3.02E-03</u>	4.32E-03	6.87E-03
$GD(\sigma)$	<u>2.74E-04</u>	3.82E-04	5.56E-04

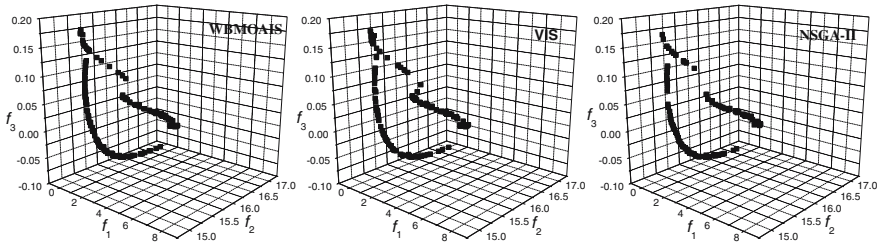


Fig. 3. Pareto-optimal front obtained using WBMOAIS, VIS and NSGA-II for VNT

5 Conclusion

In this study, based on opt-aiNET, a novel multiobjective artificial immune algorithm is presented. The proposed algorithm WBMOAIS is compared with VIS and NSGA-II. Numerical results on three standard problems (ZDT6, POL, VNT) show WBMOAIS has better behavior than VIS and NSGA-II for all metrics except that NSGA-II has a small value for the standard deviation of the S metric. Therefore, it can be concluded that WBMOAIS overall outperforms VIS and NSGA-II for all test problems.

References

1. Deb, K.: Multi-objective Optimization Using Evolutionary Algorithms. John Wiley & Sons, Ltd., New York (2001)
2. Coello Coello, C.A., Lamont, G.B., Van Veldhuizen, D.A.: Evolutionary Algorithms for Solving Multi-objective Problems, 2nd edn. Springer Science, New York (2008)
3. de Castro, L.N., Timmis, J.: Artificial Immune System: A New Computational Intelligence Approach. Springer, Heidelberg (2002)
4. Hart, E., Timmis, J.: Application Areas of AIS: The Past, the Present and the Future. Appl. Soft. Comput. 3, 191–201 (2008)
5. Smith, R.E., Forrest, S., Perelson, A.S.: Population Diversity in an Immune System Model: Implication for Genetic Search. In: Darrel Whitley, L. (ed.) Foundation of Genetic Algorithm 2, pp. 153–165. Morgan Kaufmann, San Mateo (1993)
6. Kurpati, A., Azarm, S.: Immune Network Simulation with Multiobjective Genetic Algorithms for Multidisciplinary Design Optimization. Eng. Optimiz. 33, 245–260 (2000)
7. Yoo, J., Hajela, P.: Immune Network Simulations in Multicriterion Design. Struct. Optimiz. 18, 85–94 (1999)
8. Coello Coello, C.A., Cruz Cortés, N.: An Approach to Solve Multiobjective Optimization Problems Based on an Artificial Immune System. In: Timmis, J., Bentley, P.J. (eds.) First International Conference on Artificial Immune Systems (ICARIS 2002), pp. 212–221. University of Kent, Canterbury (2002)
9. Coello Coello, C.A., Cruz Cortés, N.: Solving Multiobjective Optimization Problems Using an Artificial Immune System. Genet. Prog. Evol. Mach. 6, 163–190 (2005)

10. de Castro, L.N., Timmis, J.: An Artificial Immune Network for Multimodal Function Optimization. In: Proc. 2002 Congress on Evolutionary Computation, CEC 2002, Honolulu, vol. 1, pp. 699–704. IEEE Press, Los Alamitos (2002)
11. Freschi, F., Repetto, M.: VIS: An Artificial Immune Network for Multi-objective Optimization. *Eng. Optimiz.* 38(8), 975–996 (2006)
12. Luh, G.C., Chueh, C.H., Liu, W.W.: Multi-objective Optimal Design of Truss Structure with Immune Algorithm. *Comput. Struct.* 82, 829–844 (2004)
13. Jiao, L., Gong, M., Shang, R., Du, H., Lu, B.: Clonal Selection with Immune Dominance and Energy Based Multiobjective Optimization. In: Coello Coello, C.A., Hernández Aguirre, A., Zitzler, E. (eds.) EMO 2005. LNCS, vol. 3410, pp. 474–489. Springer, Heidelberg (2005)
14. Zhang, X., Lu, B., Gou, S., Jiao, L.: Immune Multiobjective Optimization Algorithm for Unsupervised Feature Selection. In: Rothlauf, F., Branke, J., Cagnoni, S., Costa, E., Cotta, C., Drechsler, R., Lutton, E., Machado, P., Moore, J.H., Romero, J., Smith, G.D., Squillero, G., Takagi, H. (eds.) EvoWorkshops 2006. LNCS, vol. 3907, pp. 484–494. Springer, Heidelberg (2006)
15. Wang, X.L., Mahfouf, M.: ACSAMO: An Adaptive Multiobjective Optimization Algorithm Using the Clonal Selection Principle. In: Proc. 2nd European Symposium on Nature-inspired Smart Information Systems, Puerto de la Cruz, Tenerife, Spain (2006)
16. Zhang, Z.H.: Multiobjective Optimization Immune Algorithm in Dynamic Environments and Its Application to Greenhouse Control. *Appl. Soft. Comput.* 8, 959–971 (2008)
17. Deb, K., Pratap, A., Agarwal, S., et al.: A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* 6(2), 182–197 (2002)

A Neural Network Model for Solving Nonlinear Optimization Problems with Real-Time Applications

Alaeddin Malek and Maryam Yashtini

Department of Mathematics, Tarbiat Modares University,
Tehran 14115-175, Iran
{mala,m_yashtini}@modares.ac.ir

Abstract. A new neural network model is proposed for solving nonlinear optimization problems with a general form of linear constraints. Linear constraints, which may include equality, inequality and bound constraints, are considered to cover the need for engineering applications. By employing this new model in image fusion algorithm, an optimal fusion vector is exploited to enhance the quality of fused images efficiently. The stability and convergence analysis of the novel model are proved in details. The simulation examples are used to demonstrate the validity of the proposed model.

Keywords: Neural network model, Optimization, Real-time applications, Stability analysis.

1 Introduction

Neural networks have been extensively studied over the past few decades and have found applications in variety of areas such as associative memory, moving object speed detection, image and signal processing, and pattern recognition. In many applications, real-time solutions are usually imperative [1,2,3]. These applications strongly depend on the dynamic behavior of the networks. In the recent years, due to the in-depth research in neural networks, numerous dynamic solvers based on neural networks have been developed and investigated [4], [5], [6-17]. Specially, in the past two decades, various neural network models have been developed for solving the linearly constrained nonlinear optimization problems, e.g., those based on the penalty parameter method [7], the Lagrange method [9], the gradient and projected method [5], [13], the primal-dual method [12], [14], and the dual method [4], [15]. Malek and his coauthors in reference [19] presented a recurrent neural network for solving linear and quadratic optimization problems. Their network is shown to be globally convergent to an exact optimal solution of the linear or quadratic optimization problems. Their network is not suitable for solving nonlinear optimization problems. In this paper, we propose a one-layer neural network for solving nonlinear optimization problems with general linear constraints. In particular, since simple structure and global stability are the most desirable dynamic properties of the neural networks, our motivation of this study is mainly focused on developing a neural network with these properties adequate for solving nonlinear real-time optimization problems. Another objective of this paper is to concern with the real

time application of the proposed neural network model in image fusion algorithm to restore noisy images. Theoretical aspects and the illustrative examples further show the effectiveness and applicability of the proposed model.

2 Artificial Neural Network Formulation

In this section, we describe the nonlinear programming problem and discuss its equivalent formulation. Then we will propose a recurrent neural network to solve this problem.

Consider the following nonlinear programming problem subject to general linear constraints:

$$\text{Minimize } f(x) \text{ subject to } Bx = b, Ax \leq d, \ell \leq x \leq h, \quad (1)$$

where $f(x)$ is a continuously differentiable and convex from R^n to R , $B \in R^{m \times n}$, $A \in R^{r \times n}$, $x, \ell, h \in R^n$, $b \in R^m$ and $d \in R^r$.

According to the Karush-Kuhn-Tucker (KKT) conditions and well known projection theorem [20], we see that x^* is optimal solution for problem (1) if and only if there exist $y^* \in R^r$ and $w^* \in R^m$ such that $((x^*)^T, (y^*)^T, (w^*)^T)^T$ satisfies the following conditions:

$$\begin{cases} x = g_1(x - \alpha(\nabla f(x) + A^T y - B^T w)) \\ y = g_2(y + \alpha Ax - \alpha d) \\ Bx = b, \end{cases} \quad (2)$$

where α is positive constant, $g_2(y) = [g_2(y_1), \dots, g_2(y_m)]^T$, in which $g_2(y_i) = \max\{0, y_i\}$, $g_1(x) = [g_1(x_1), \dots, g_1(x_n)]^T$, where for $i = 1, \dots, n$

$$g_1(x_i) = \begin{cases} \ell_i & x_i < \ell_i \\ x_i & \ell_i \leq x_i \leq h_i \\ h_i & x_i > h_i. \end{cases}$$

The Eq. (2) can be equivalently written as

$$\begin{pmatrix} x \\ y \\ w \end{pmatrix} = \begin{pmatrix} g_1(x - \alpha(\nabla f(x) + A^T y - B^T w)) \\ g_2(y + \alpha Ax - \alpha d) \\ w - \alpha(Bx - b) \end{pmatrix}. \quad (3)$$

Based on Eq. (3), we propose a artificial neural network for solving problem (1), with the following dynamical equation:

$$\frac{dz}{dt} = \frac{d}{dt} \begin{pmatrix} x \\ y \\ w \end{pmatrix} = \lambda \begin{pmatrix} -x + g_1(x - \alpha(\nabla f(x) + A^T y - B^T w)) \\ -y + g_2(y + \alpha Ax - \alpha d) \\ -\alpha(Bx - b) \end{pmatrix} = \lambda H(z). \quad (4)$$

where $\lambda > 0$ is a positive constant and $z = (x^T, y^T, w^T)^T \in R^{n+m+r}$ is a state vector. The simplified architecture of the artificial neural network (4) is shown in Fig. 1.

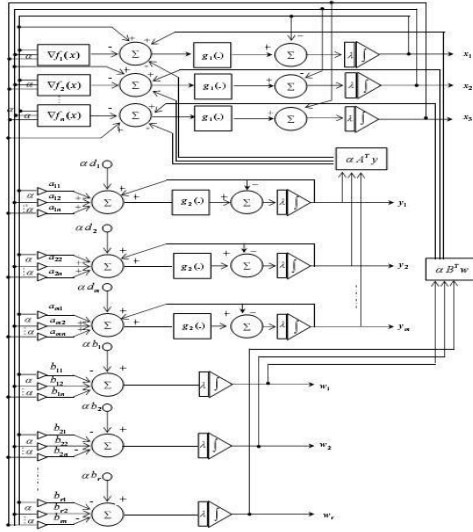


Fig .1. Depict of simplified architecture of the artificial neural network (4)

3 Theoretical Aspects

In this section, we will study some properties of the artificial neural network (4). Through this section we assume that $\nabla^2 f(x)$ is positive definite on $X = \{x \in \mathbb{R}^n \mid \ell \leq x \leq \bar{h}\}$.

Lemma 1. *For any initial point $z^0 = ((x^0)^T, (y^0)^T, (w^0)^T)^T \in X \times \mathbb{R}_+^m \times \mathbb{R}^r$, there exists a unique continuous solution $z(t) = (x(t)^T, y(t)^T, w(t)^T)^T$ for the artificial neural network (4).*

Proof. Since the projection operators g_1 and g_2 are locally Lipschitz continuous, $H(z)$ is also locally Lipschitz continuous. Thus, According to the local existence and uniqueness theorem of ordinary differential equations [21], there exists a unique continuous solution $z(t) = (x(t), y(t), w(t))^T$ for (t_0, T) . We will show that $z(t)$ is bounded and the local existence for solution of (4) can be extended to global existence.

Theorem 1. *Let $z(t)$ be the state trajectory of (4) with the initial point $z^0 \in X \times \mathbb{R}_+^m \times \mathbb{R}^r$. Then the proposed artificial neural network of (4) is stable in the Lyapunov sense and globally convergent to the stationary point $z^* = (x^*, y^*, w^*)$, where x^* is the optimal solution of problem (1).*

Proof. We define the following Energy function:

$$V(z, z^*) = -W(z)^T H(z) - \frac{1}{2} \|H(z)\|^2 + \frac{1}{2} \|z - z^*\|^2$$

where

$$W(z) = \begin{pmatrix} \nabla f(x) + A^T y - B^T w \\ -Ax + d \\ Bx - b \end{pmatrix}$$

z^* is an equilibrium point of artificial neural network (4).

Let $\hat{S} \subseteq R^{n+m+r}$ be a neighborhood of z^* . We show that $V(z, z^*)$ is a Lyapunov function proper to neural network (4). By the results give in [22], we know that

$$-W(z)^T H(z) \geq \|H(z)\|^2 \tag{5}$$

$$(H(z) + z - z^*)^T (-H(z) - W(z)) \geq 0. \tag{6}$$

It is obvious that $V(z, z^*) \geq \frac{1}{2} \|z - z^*\|^2$, and for all $z \in \hat{S} \setminus \{z^*\}$, $V(z, z^*) > 0$.

In the following, we show that $\frac{dV(z, z^*)}{dt} \leq 0$. Since $\frac{dV}{dt} = \nabla V(z(t), z^*)^T \frac{dz(t)}{dt}$, then from theorem 3.2 of [23], we know that

$$\nabla V(z, z^*) = W(z) - (\nabla W(z) - I)H(z) + z - z^*$$

where $\nabla W(z)$ denotes the Jacobian matrix of W . Then

$$\begin{aligned} \frac{dV(z, z^*)}{dt} &= (W(z) - (\nabla W(z) - I)H(z) + z - z^*)^T H(z), \\ &= (W(z) + z - z^*)^T H(z) + \|H(z)\|^2 - H(z)^T \nabla W(z)H(z). \end{aligned}$$

From (5) we can write $(W(z) + z - z^*)^T H(z) \leq -(z - z^*)^T W(z) - \|H(z)\|^2$. Thus

$$\frac{dV(z, z^*)}{dt} \leq -(z - z^*)^T W(z) - H(z)^T \nabla W(z)H(z) \tag{7}$$

Since $\nabla W(z)$ is a positive semi-definite matrix, $(z - z^*)^T W(z) \geq 0$ and $H(z)^T \nabla W(z)H(z) \geq 0$. So

$$\frac{dV(z, z^*)}{dt} \leq -(z - z^*)^T W(z) - H(z)^T \nabla W(z)H(z) \leq 0. \tag{8}$$

Therefore, the function $V(z, z^*)$ is an Energy function suitable for (4). From (8), $V(z, z^*)$ is monotonically nonincreasing for all $t \geq t_0$.

It is easy to see that $\phi = \{z \in R^{n+m+r} \mid V(z, z^*) \leq V(z^0, z^*)\}$ is bounded since

$$V(z^0, z^*) \geq V(z, z^*) \geq \frac{1}{2} \|H(z)\|^2 + \frac{1}{2} \|z - z^*\|^2 \geq \frac{1}{2} \|z - z^*\|^2 \geq 0,$$

therefore $T = \infty$.

Because $V(z, z^*)$ is radially bounded, for any initial point $z^0 \in X \times R_+^m \times R^r$, there exists a convergent subsequence $\{z(t_k)\}$ such that $\lim_{k \rightarrow \infty} z(t_k) = \hat{z}$, where

$$\frac{dV(\hat{z}, z^*)}{dt} = 0. \text{ It can be seen that } dV(\hat{z}, z^*)/dt = 0 \text{ implies}$$

$$(\hat{z} - z^*)^T W(\hat{z}) + H(\hat{z})^T \nabla W(\hat{z}) H(\hat{z}) = 0. \tag{9}$$

That is,

$$H(\hat{z})^T \nabla W(\hat{z}) H(\hat{z}) = 0, \quad (z - z^*)^T W(\hat{z}) \geq 0, \quad (W(\hat{z}) - W(z^*))^T (\hat{z} - z^*) = 0. \tag{10}$$

Let $\hat{z} = (\hat{x}, \hat{y}, \hat{w}) \in R^{n+m+r}$. Then equality $H(\hat{z})^T \nabla W(\hat{z}) H(\hat{z}) = 0$ implies that

$$[g_1(\hat{x} - \nabla f(\hat{x}) - A^T \hat{y} + B^T \hat{z}) - \hat{x}]^T \nabla^2 f(\hat{x}) \times [g_1(\hat{x} - \nabla f(\hat{x}) - A^T \hat{y} + B^T \hat{z}) - \hat{x}] = 0.$$

The positive-definiteness of $\nabla^2 f(\hat{x})$ implies that

$$[g_1(\hat{x} - F(\hat{x}) - A^T \hat{y} + B^T \hat{z}) - \hat{x}] = 0. \tag{11}$$

Now, from $(W(\hat{z}) - W(z^*))^T (\hat{z} - z^*) = 0$ we can write

$$(\nabla f(\hat{x}) - \nabla f(x^*))^T (\hat{x} - x^*) = (x - x^*)^T \nabla^2 f(x_\mu) (\hat{x} - x^*) = 0$$

where $x_\mu = (1 - \mu)\hat{x} + \mu x^*$ for all $0 \leq \mu \leq 1$. It follows that $\hat{x} = x^*$, thus

$$B\hat{x} - b = 0. \tag{12}$$

From $(z - z^*)^T W(\hat{z}) \geq 0$ we can get

$$(\hat{x} - x^*)^T (\nabla f(\hat{x}) - A^T \hat{y} - B^T \hat{w}) + (\hat{y} - y^*)^T (A\hat{x} - d) + (\hat{w} - w^*)^T (B\hat{x} - b) = 0.$$

Since $\hat{x} = x^*$, it is equivalently written as $(\hat{y} - y^*)^T (-A\hat{x} + d) = 0$. Then

$$\hat{y}^T (-A\hat{x} + d) = (y^*)^T (-A\hat{x} + d) = (y^*)^T (-Ax^* + d) = 0.$$

Furthermore $\hat{y}^T (-A\hat{x} + d) = 0$, $\hat{y} \geq 0$ and $-A\hat{x} + d \geq 0$ if and only if

$$g_2(\hat{y} + A\hat{x} - b) - \hat{y} = 0. \tag{13}$$

Thus from Eqs. (11)-(13), the point $\hat{z} = (\hat{x}, \hat{y}, \hat{w})$ satisfies in Eq. (2). This means \hat{z} is an equilibrium point of artificial neural network (4).

Now we consider another function

$$\hat{V}(z, \hat{z}) = -W(z)^T H(z) - \frac{1}{2} \|H(z)\|^2 + \frac{1}{2} \|z - \hat{z}\|^2$$

where $\hat{z} = (\hat{x}, \hat{y}, \hat{w})$. Similar to the previous analysis, we have $\frac{d\hat{V}(z, \hat{z})}{dt} \leq 0$ and

$\lim_{k \rightarrow \infty} \hat{V}(z(t_k), \hat{z}) = 0$. So, for $\forall \varepsilon > 0$ there exists $q > 0$ such that when $t_k \geq t_q$ we have $\hat{V}(z(t_k), \hat{z}) < \varepsilon^2 / 2$. Since $\hat{V}(z(t), \hat{z})$ decreases as $t \rightarrow \infty$ for $t \geq t_q$, $\|z(t) - \hat{z}\| \leq \sqrt{2\hat{V}(z(t), \hat{z})} \leq \sqrt{2\hat{V}(z(t_k), \hat{z})} < \varepsilon$. Then $\lim_{t \rightarrow \infty} z(t) = \hat{z}$.

Therefore, the state trajectory of the proposed neural network is globally convergent to an equilibrium point of (4).

4 Real-Time Application

The proposed artificial neural network in (4) can be applied for many real-time optimization problems. For example, this model can be used to support vector machines for classification and regression and to robot motion control in real-time. Here, we apply the proposed artificial neural network (4) to increase the useful information content of images and improve the quality of the noisy images. The neural fusion algorithms are proposed in details for noisy images.

Consider an array of n sensor. Let $I_l(k)$ denote the received two dimensional images with $M \times N$ gray-level from the l th sensor. Let its amplitude is denoted by $f_l(i, j)$, which

$$I_l((i-1)N + j) = f_l(i, j), \quad (i = 1, \dots, M; j = 1, \dots, N). \quad (14)$$

The images consist of the desired image $s(k)$, scaling coefficient a_l , and the measured noise $\hat{n}_l(k)$. Then the n – dimensional vector of information received from n sensors is given by $I(k) = a s(k) + \hat{n}(k)$, where

$$a = [a_1, \dots, a_n]^T, I(k) = [I_1(k), \dots, I_n(k)]^T, \text{ and } \hat{n}(k) = [\hat{n}_1(k), \dots, \hat{n}_n(k)]^T.$$

According to the result discussed in [18], the considered image fusion problem can be formulated as a deterministic quadratic programming problem

$$\min \quad f(w) = x^T R x \quad \text{subject to} \quad a^T x = 1, x \geq 0 \quad (15)$$

where $a = [1, \dots, 1]^T \in \mathfrak{R}^n$, $R = \frac{1}{MN} \sum_{k=1}^{MN} I(k)I(k)^T$ (sample variance matrix) and

$x = [x_1, \dots, x_n]^T$ is called the fusion vector. Based on the proposed artificial neural network (4) for solving (1), we propose the following artificial neural network for solving (15), which its state equation

$$\frac{d}{dt} \begin{pmatrix} x \\ w \end{pmatrix} = \begin{pmatrix} -x + g_1(x - Rx + a^T w) \\ -a^T x + 1 \end{pmatrix}. \quad (16)$$

The artificial neural network (16) converges to optimal fusion vector x^* . The output equation is

$$s^*(k) = \sum_{l=1}^n x_l^* I_l(k), \quad (17)$$

After increasing the useful information content of images by using Eq. (17), it is time to see the fused image. First we have to convert s^* to f^* as follows

$$f^*(i, j) = s^*((i-1)N + j), \quad (i = 1, \dots, M; j = 1, \dots, N)$$

then use the function *imshow*(f^*) in Matlab.

One can be improve the quality of fused images using the proposed neural image fusion algorithm by increasing the number of sensors.

Theorem 2. *Let R be a positive definite matrix, then the artificial neural network defined in (16) converges to optimal fusion vector $(x^*, w^*) \in \mathfrak{R}^{K+1}$.*

5 Simulation Examples

In this section, two examples are provided to illustrate both the theoretical results achieved in Sections 3 and 4 and the simulation performance of artificial neural networks (4) and (16). The simulations are conducted in matlab and 4th order Runge-Kutta technique is used for implementation.

Example 1. Consider the following nonlinear programming problem

$$\begin{aligned} &\text{Minimize } 1.05x_1^2 + x_2^2 + x_3^2 + x_4^2 - 4x_1x_2 - 2x_2 - x_4, \\ &\text{Subject to } \begin{cases} 2x_1 + x_2 + x_3 + 4x_4 = 7 \\ 2x_1 + 2x_2 + 2x_4 = 6, \\ x_1 \geq 1.5, x_2 \geq 0.5, x_3 \geq 1.5, x_4 \geq 1. \end{cases} \end{aligned} \tag{18}$$

This problem has a unique optimal solution $x^* = [2.5, 0, 0, 0.5]^T$. We solve (18) by using the proposed artificial neural network (4). The simulation results show that the proposed neural network (4) is always convergent to an equilibrium point of (18) and its output trajectory is always exponentially convergent to x^* . For example, let $\lambda=10$, Figs 2 and 3 display the convergence of behavior $z(t)$ and $\|x(t) - x^*\|^2$ based on feasible primal and arbitrary dual initial point. The convergence of behavior $z(t)$ and $\|x(t) - x^*\|^2$ is displayed based on infeasible primal and arbitrary dual initial point in Figs 4 and 5.

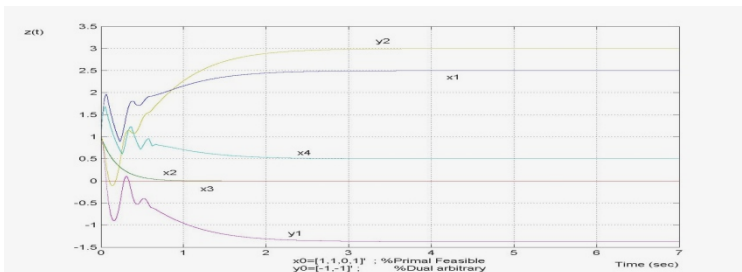


Fig. 2. Example 1: Display of transient behavior of the artificial neural network (4) with initial point $z^0 = (1, 1, 0, 1, -1, -1)^T$

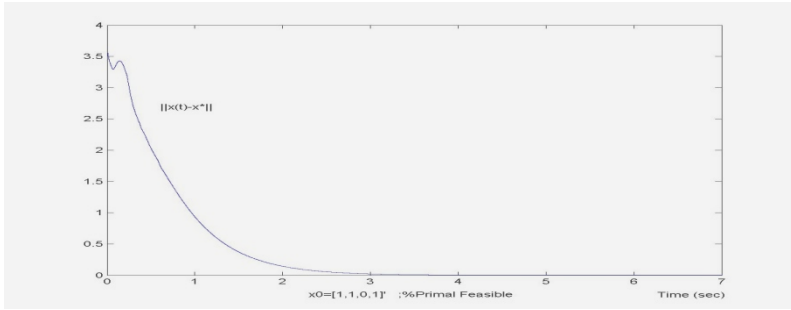


Fig. 3. Example 1: Display of transient behavior of the norm $\|x(t) - x^*\|^2$ based on artificial neural network (4) with initial point $z^0 = (1, 1, 0, 1, -1, -1)^T$

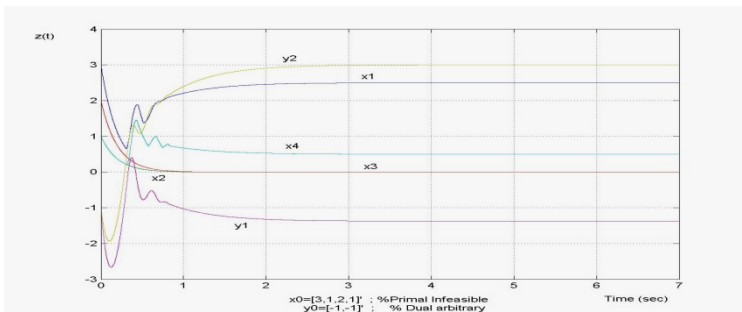


Fig. 4. Example 1: Display of transient behavior of the artificial neural network (4) with initial point $z^0 = (3, 1, 2, 1, -1, -1)^T$

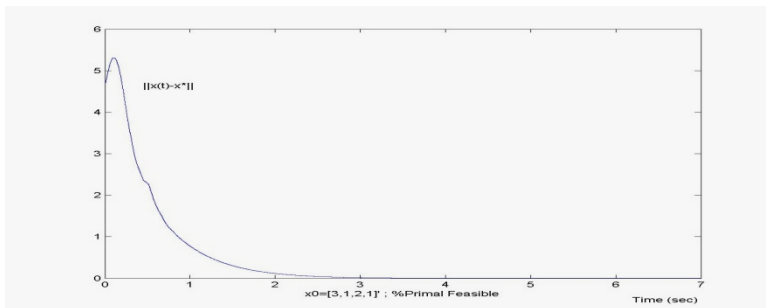


Fig. 5. Example 1: Display of transient behavior of the norm $\|x(t) - x^*\|^2$ based on artificial neural network (4) with initial point $z^0 = (3, 1, 2, 1, -1, -1)^T$

Example 2. This example investigates the performance of the proposed artificial neural network (16) to a neural image fusion algorithm. The proposed neural image fusion algorithm is applied to the Lena image shown in Fig. 6. It is an eight-bit

gray-level image with 206 by 245 pixels. Fig. 6(a) is a noisy Lena image measured by one sensor, where its SNR is 9dB. Figs. 6(b)-(d) are fused images by the proposed algorithm for the number of sensors $n=10, 20$ and 30 , respectively. Apparently, the quality of the fused image shown in Fig. 6(b)-(d) is improving as n increases. Table 1 shows that with a fix number of iterations, the quality of images improve when the number of sensor increases.

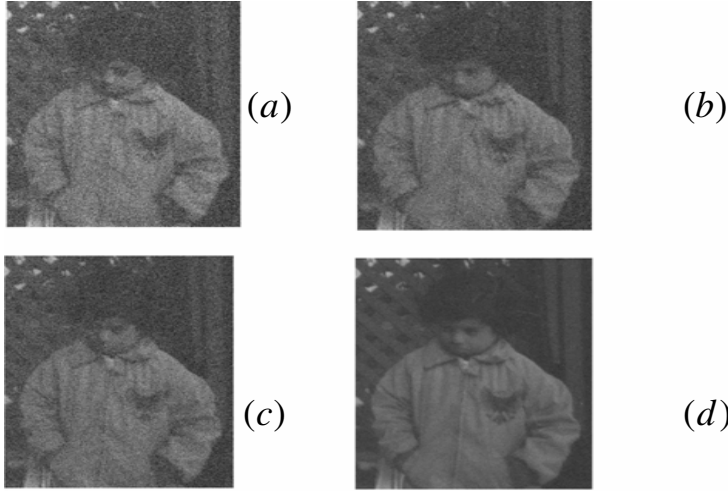


Fig. 6. Example 2: Lena Image fusion using the neural fusion algorithm (a) The noisy image. (b)-(d) The fused image with $n=10, 20, 30$.

Table 1. Display of CPU times and the error $\|s(k) - s^*(k)\|_2$ for gray level fused images with 10, 20 and 30 sensors. A fixed number of iterations are used in neural image fusion algorithm.

Number of sensors	Number of iterations	CPU time	2-norm error $\ s(k) - s^*(k)\ _2$
10	7	23.5340	6.1297×10^{-4}
20	7	37.3010	4.0075×10^{-4}
30	7	43.0930	2.5700×10^{-4}

6 Concluding Remarks

We have proposed a recurrent neural network model for solving constrained nonlinear optimization problems. We have proved that the proposed model is Lyapunov stable and converges globally to unique optimal solution of the constrained nonlinear optimization problems. Moreover, the new model is amenable to parallel implementation.

We have proposed a neural fusion algorithm based on the proposed neural network model. This algorithm is simple for implementation. Numerical results show a good agreement with theoretical aspects.

References

1. Kalouptisidis, N.: *Signal Processing Systems, Theory and design*. Wiley, New York (1997)
2. Youshikawa, T.: *Foundations of Robotics: Analysis and Control*. MIT Press, Cambridge (1990)
3. Cichocki, A., Unbehauen, R.: *Neural Networks for Optimization and Signal Processing*. Wiley, England (1993)
4. Xia, X., Wang, J.: A Dual Neural Network for Kinematic Control of Redundant Robot Manipulators. *IEEE Trans. Systems Man, Cybernet.* 31, 147–154 (2001)
5. Bouzerdoum, A., Pattison, T.R.: Neural Network for Quadratic Optimization with Bound Constraints. *IEEE Trans. Neural Networks* 4, 293–304 (1993)
6. Tank, D.W., Hopfield, J.J.: Simple Neural Network Optimization Networks: an A/D Converter Signal Decision Network, and Linear Programming Circuit. *IEEE Trans. Circ. Syst.* 33, 533–542 (1986)
7. Kennedy, M.P., Chua, L.O.: Neural Networks for Nonlinear Programming. *IEEE. Trans. Circuits Systems* 35, 554–562 (1998)
8. Xia, Y., Wang, J.: Recurrent Neural Networks for Optimization: the State of Art. In: Medsker, L.R., Jain, L.C. (eds.), vol. 4. CRC Press, New York (2000)
9. Zhang, S., Constantinides, A.G.: Lagrange programming neural networks. *IEEE Trans. Circuits Systems* 39, 441–452 (1992)
10. Sudharsanan, S., Sundareshan, M.: Exponential Stability and Systematic Synthesis of a Neural Network for Quadratic Minimization. *Neural Networks* 4, 599–613 (1991)
11. Wang, J.: A Deterministic Annealing Neural Network for Convex Programming. *Neural Networks* 7, 629–641 (1994)
12. Xia, Y.: A New Neural Network for Solving Linear and Quadratic Programming Problems. *IEEE Trans. Neural Networks* 7, 1544–1547 (1996)
13. Liang, X.B., Wang, J.: A Recurrent Neural Network for Nonlinear Optimization with a Continuously Differentiable Objective Function and Bounded Constraints. *IEEE Trans. Neural Networks* 11, 1251–1262 (2000)
14. Tao, Q., Cao, G.D., Xue, M.S., Qiao, H.: A High Performance Neural Network for Solving Nonlinear Programming Problems with Hybrid Constraints. *Phys. Lett. A* 288, 88–94 (2001)
15. Xia, Y., Wang, J.: A General Methodology for Designing Globally Convergent Optimization Neural Networks. *IEEE Trans. Neural Networks* 9, 1331–1343 (1998)
16. Yashtini, M., Malek, A.: Solving Complementarity and Variational Inequalities Problems Using Neural Networks. *Appl. Math. Comput.* 190, 216–230 (2007)
17. Yashtini, M., Malek, A.: A Discrete-time Neural Network for Solving Nonlinear Convex Problems with Hybrid Constraints. *Appl. Math. Comput.* 195, 576–584 (2008)
18. Malek, A., Yashtini, M.: Image Fusion Algorithms for Color and Gray Level Images Based on LCLS Method and Novel Artificial Neural Network. *Neurocomputing* (submitted)

19. Oskoei, H.G., Malek, A., Ahmadi, A.: Novel Artificial Neural Network with Simulation Aspects for Solving Linear and Quadratic Programming Problems. *Comp. Math. Appl.* 53, 1439–1454 (2007)
20. Bertsekas, D.P., Tsitsiklis, J.N.: *Parallel and Distributed Computation, Numerical Methods*. Prentice-Hall, Englewood Cliffs (1989)
21. Miller, R.K., Michel, A.N.: *Ordinary Differential Equations*. In: *Neural Network*. Academic, San Diego (1982)
22. Pang, J.S.: A Posteriori Error Bounds for the Linearly-constrained Variational Inequality Problem. *Math. Oper. Res.* 12, 474–484 (1987)

Evolutionary Markov Games Based on Neural Network

Liu Weibing¹, Wang Xianjia², and Huang Binbin³

¹ School of Political Science and Public Management, Wuhan University,
Wuhan 430072, China

² Economics and Management School of Wuhan University, Wuhan 430072, China

³ Institute of system engineering, Wuhan University, Wuhan 430072, China

Abstract. Based on the dynamic characteristics of evolutionary game and Markov process, this paper presents a dynamic decision model for evolutionary Markov games. In this model, players' strategy-choosing is mapped to a Markov decision process with payoffs, and transition probability is made by Boltzmann distribution. This paper uses neural network to simulate strategy-choosing in evolutionary Markov games. Experimental results show that the neural network can successfully simulate players' dynamic learning and actions in evolutionary Markov games.

Keywords: Evolutionary game, Markov process, Decision, Neural network.

1 Introduction

Game theory, which is the subject about decision problem among multi-players, is the research on problems of competition and cooperation. In 1970's the fundamental notion of evolutionary game theory was introduced by Maynard Smith. In his book *Evolution and Theory of Games* [1] he presented an evolutionary approach in classical game theory. Evolutionary game theory has extensive applications in many fields, such as mathematics, biology, ecology, economics and sociology.

A number of publications on the building of models and analytic methods for evolutionary games have been seen during last decade. Yao and Darwen presented a method which introduced genetic algorithm into evolutionary games [2]. Thlol and Acan investigated an ant colony optimization approach for iterated prisoner's dilemma [3]. This method provides game strategies of better quality than genetic algorithms, but needs longer running times. Amir et al. proposed a dynamic model for 2×2 symmetric games using birth and death process [4]. Tadj and Touzene extended the work of M. Amir et al. and gave a QBD approach for evolutionary games [5].

Neural network (NN) is generally the software systems which imitates the networks of the human brain. And they can be applied successfully in learning, generalization, and optimization functions [6, 7]. The neural network can provide a useful methodology for the autonomous mobile robot learning. NN-based techniques are extensively used in multi-agent robot system [8, 9]. Jolly et al. used an NN-based method for intelligent decision-making in a robot soccer system [10]. The potential of NN-based techniques have not been fully utilized for multi-agent decision-making problems. In this paper, we apply neural network in our system. We simulate the player's

decision-making with neural network for learning. In evolutionary Markov games, the inputs of the neural network are decided by the last actions of players. The output is the next action that the player will choose. Simulation results show that neural network can successfully simulate evolutionary Markov games.

2 Some Preliminaries

2.1 Markov Process

Supposing $I=\{0,1,2,\dots\}$ is the space of state, $T=\{0,1,2,\dots\}$ is a set of time. A stochastic process $\{X_t, t \in T\}$ is called Markov if for arbitrary t and state i_0, i_1, \dots, i_n, j , we have

$$\begin{aligned} P\{X_{t+1} = j \mid X_t = i, X_{t-1} = i_{n-1}, \dots, X(1) = i_1, X(0) = i_0\} \\ = P\{X_{t+1} = j \mid X_t = i\} \end{aligned}$$

That is to say, the Markov process whose future probabilities are determined by its most recent values.

The conditional probability

$$P_{ij}(1) = P\{X_{t+1} = j \mid X_t = i\}$$

is called one-step transition probability. It means the probability of state transition from i to j , so $P_{ij}(1)$ satisfies the following:

- (1) $P_{ij}(1) \geq 0 \quad i, j \in I$
- (2) $\sum P_{ij}(1) = 1 \quad i \in I$

Therefore, we have the one-step Markov transition probability matrix

$$P_1 = \begin{bmatrix} P_{00}(1) & P_{01}(1) & \dots & \dots \\ \dots & \dots & \dots & \dots \\ P_{n0}(1) & P_{n1}(1) & \dots & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}$$

2.2 Game Theory

We restrict our view to the class of finite games in strategic form. Generally a normal form game consists of three key components:

- (1) Players: Let $I=\{1,2,\dots,n\}$ be a set of players, where n is a positive integer.
- (2) Strategies space: For each player $i \in I$, let S_i denote a set of allowable actions, called the pure strategies set. The choice of a specific action $s_i \in S_i$ of a player $i \in I$ is called a pure strategy. The vector $s=\{s_1, s_2, \dots, s_n\}$ is called a pure strategies profile.
- (3) Payoff function: For any strategies profile s and any player $i \in I$, let $u_i(s_1, s_2, \dots, s_n)$ be the payoff to player i .

Evolutionary game theory is the extension of classic game theory, and evolutionary games are dynamic games.

2.3 Neural Network

The neural network is used to make the internal models. Building the model of an agent's behavior may involve many methods such as finite automata, Markov chain, and so on. Neural network is generally the software systems that imitate the networks of the human brain. And they can be applied successfully in learning, relating, classification, generalization, and optimization functions. It appears that the multi-layer neural network can be trained to approximate and accurately generalize virtually and smooth, measurable functions. Because neural network has the ability to work with incomplete data, they can easily form models for many complex problems.

The neural network is consisted of many units that represent neurons. Each unit is a basic unit of information process. Units are interconnected via links that contain weight values. Weight values help the neural network to express knowledge. Figure 1 is the general topology of a neural network.

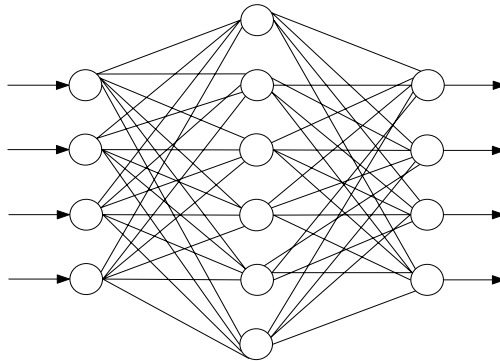


Fig. 1. The topology of a 3-layer neural network

3 Evolutionary Markov Games

We restrict our view to the repeated games. In repeated games, players update strategies by their payoffs. In this paper, we looked iteration times of games as a set of time $t=1,2,\dots$ in Markov process $\{X(t), t \in T\}$, strategy space was mapped to the state space I of Markov process, and the payoffs in repeated games were seemed as rewards for state transition. Therefore, the repeated games were mapped to a Markov process with rewards. For example, the iterated prisoner's dilemma assumes a binary choice for the players, either cooperation (C) or defection (D). So the iterated prisoner's dilemma was looked as a Markov process with two-state transition. More generally, if at time t (the t -th repeated game) the player is in state i ($i=C,D$). Then at time $t+1$ the probability that the player chooses strategy C or D can be obtained by Markov transition probability matrix (Fig. 2). Where $p_{ij}=P\{X_{t+1}=j | X_t=i\}$, $i, j=C$ or D .

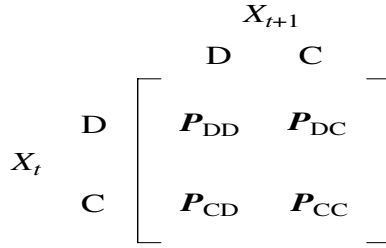


Fig. 2. Markov process for iterated prisoner’s dilemma

As a Markov process with N states, supposing u_{ij} is the payoff when the state transferred from i to j , u_{ij} can be also seemed as a reward for state transition. With the transition of states, the system will generate a series of rewards. So that the rewards u_{ij} are stochastic variables, and its probability distribution is determined by Markov transition probability matrix.

Consider 2-player symmetric repeated games, supposing now the player is in state i (namely the player’s strategy is i), we can get the total expected payoff after n times repeated games using Markov process. Firstly, we assume $v_i(t)$ is the total expected payoff after t times transition from state i . We can get the following equation:

$$v_i(t) = \sum_{j=1}^N p_{ij} [u_{ij} + v_j(t-1)] \quad , (i = 1, 2, \dots, N; t = 1, 2, 3, \dots) \quad (1)$$

equation (1) can be simplified as follows

$$v_i(t) = \sum_{j=1}^N p_{ij} u_{ij} + \sum_{j=1}^N p_{ij} v_j(t-1) \quad (2)$$

If define $q_i = \sum_{j=1}^N p_{ij} u_{ij} \quad (i = 1, 2, \dots, N) \quad (3)$

q_i can be explained as the expected reward after one time transition from state i . So that we can obtain

$$v_i(t) = q_i + \sum_{j=1}^N p_{ij} v_j(t-1) \quad (4)$$

Using the form of vectors (4) will be denoted as the following:

$$V(t) = Q + PV(t-1) \quad (n = 1, 2, 3, \dots) \quad (5)$$

Where, $V(t)$ is the column vector of $v_i(t)$, Q is the column vector of q_i , and P is the Markov transition probability matrix.

From (5), if we want to obtain the total expected payoff $V(t)$, firstly we should get the Markov transition probability matrix P . In this paper, we adopt the Boltzmann

distribution. The Boltzmann distribution provides one method, where the probability of state transition from state i to j at time t is

$$p_{ij} = \frac{e^{u_j(t)/\lambda}}{\sum_k e^{u_k(t)/\lambda}} \quad (k = 1, 2, \dots, N) \tag{6}$$

Where, $u_j(t)$ is the payoff when the player chooses strategy j . The parameter λ has an important role in the learning process. By increasing λ , we can increase the randomness of decisions. On the other hand, decreasing λ will result in decreasing the randomness, this enables the player to choose the optimal strategy more accurately.

4 Examples and Numerical Results

In this section, we consider that neural networks match pair to play iterated prisoner's dilemma games. The prisoner's dilemma, is a classic model in game theory, and is a non-cooperative, non-zerosum game. Each player has two choices; either cooperation or defection. The payoffs players got are calculated according to Table 1.

Table 1. Payoffs in prisoner's dilemma

		Column player	
		Cooperation	Defection
Row player	Cooperation	3 3	0 5
	Defection	5 0	1 1

From Table 1 we can know that defection is a dominant action, so any rational player will choose defection no matter what others choose. But if they cooperate, they would get more. This is the dilemma of prisoners. To overcome this problem, Axelrod presented the thought of iterated prisoner's dilemma, which repeats the conventional game numerous times with the number of repetitions unknown to both players. Repeating the games in this way can give players the hope of cooperation. The iterated prisoner's dilemma is widely used to model systems in biology, sociology, psychology and economics. In this section, we use neural networks to play iterated prisoner's dilemma game.

Firstly we let a neural network to play iterated prisoner's dilemma games with a human. We used 3-layer neural network composed of an input layer, one hidden layer and one output layer. The input layer is composed of 6 input units, which is his and opponent's last actions. The hidden layer is composed of 20 units. Simulation parameters are showed in Table 2. Figure 3 is the evolutionary curve of iterated prisoner's dilemma games played by human and neural network.

Table 2. Simulation parameters

• Action information: 1 and 0 denote cooperation and defection respectively
• Maximal iteration times: 5000
• Minimal error of mean square: 0.01
• Number of input unit at the neural network: 6
• Number of hidden unit at neural network: 20
• Number of output unit at the neural network: 1

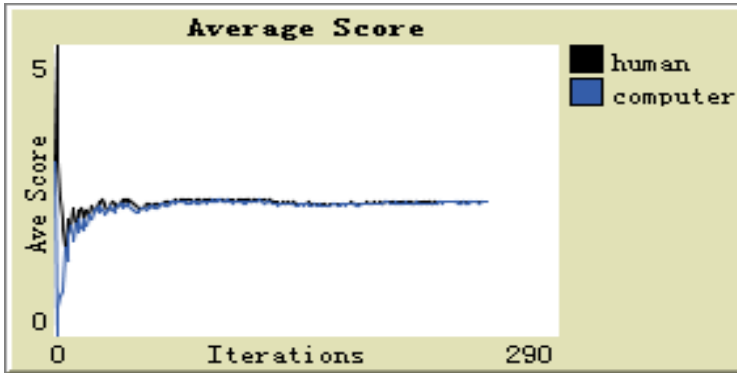


Fig. 3. The evolutionary curve of human and neural network

Secondly, we use neural network to paly with Tit-for-Tat (TFT) strategy. The TFT strategy is a very famous strategy submitted by Anatol Rapoport, starts with a move to cooperation, but thereafter repeats the last move made by its opponent. Figure 4 is the curve of epoch and error by neural network. From the results we know that after 217 times iterated games the neural network can successfully play games with TFT strategy.

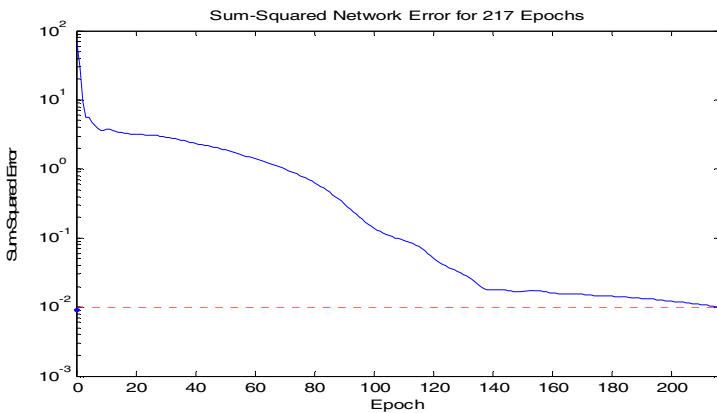


Fig. 4. The curve of epoch-error

5 Conclusions

As an extension of classical game theory, evolutionary game theory has attracted much attention in recent years. Specially, the study on learning model of players in evolutionary games is a research focus. In this paper, we introduce Markov process into evolutionary games and present a dynamic evolutionary Markov games, then use neural networks to simulate the strategy-choosing in evolutionary Markov games. In the end, simulation experiments show that neural networks can successfully be applied in evolutionary Markov games.

Acknowledgements. We would like to thank the editor and the reviews for their consideration. We also acknowledge the support by the National Nature Science Foundation of China (70533040) and the National Society Science Foundation of China (07BSH002).

References

1. Smith, J.M.: *Evolution and the Theory of Games*. Cambridge University Press, Cambridge (1982)
2. Yao, X., Darwen, P.: Genetic Algorithms and Evolutionary Games. In: Barnett, W., Chiarella, C., Keen, S., Marks, R., Schnabl, H. (eds.) *Commerce, Complexity and Evolution*, pp. 313–333. Cambridge University Press, Cambridge (2000)
3. Thlol, Y., Acan, A.: Ants Can Play Prisoner's Dilemma. In: *IEEE International Conference on Systems, Man and Cybernetics*, pp. 1348–1354 (2003)
4. Amir, M., Berninghaus, S.K.: Another Approach to Mutation and Learning. *Games and Economic Behavior* 14, 19–43 (1996)
5. Tadj, L., Touzene, A.: A QBD Approach to Evolutionary Game Theory. *Applied Mathematical Modelling* 27, 913–927 (2003)
6. Arditi, D., Oksay, F.E., Tohdemir, O.B.: Predicting the Outcome of Construction Litigation Using Neural Network. *Computer Aided Civil and Infrastructure Engineering* 13(4), 75–81 (1998)
7. Ghezzi, D., Pedrocchi, A., Menegon, A.: PhotoMEA: An Opto-Electronic Biosensor for Monitoring in Vitro Neuronal Network Activity. *Biosystems* 78(7), 150–155 (2007)
8. Kim, Y.H., Lewis, F.L.: Neural Network Output Feedback Control of Robot Manipulators. *IEEE Transactions on Robotics and Automation* 15(2), 301–309 (1999)
9. Gu, D., Hu, H.: Neural predictive control for a car-likemobile robot. *Robotics and Autonomous Systems* 39(5), 73–86 (2002)
10. Huang, H., Liang, C.: Strategy-Based Decision Making of a Soccer Robot System Using a Real-time Self-organizing Fuzzy Decision Tree. *Fuzzy Sets and Systems* 127(3), 49–64 (2002)

Another Simple Recurrent Neural Network for Quadratic and Linear Programming

Xiaolin Hu and Bo Zhang

State Key Laboratory of Intelligent Technology and Systems,
Tsinghua National Laboratory for Information Science and Technology (TNList),
Department of Computer Science and Technology,
Tsinghua University, Beijing 100084, China

Abstract. A new recurrent neural network is proposed for solving quadratic and linear programming problems, which is derived from two salient existing neural networks. One of the predecessors has lower structural complexity but were not shown to be capable of solving degenerate QP problems including LP problems while the other does not have this limitation but has higher structural complexity. The proposed model inherits the merits of both models and thus serves as a competitive alternative for solving QP and LP problems. Numerical simulations are provided to demonstrate the performance of the model and validate the theoretical results.

Keywords: Recurrent neural network, Optimization, Linear programming, Quadratic programming, Stability analysis.

1 Introduction

Consider solving the following convex quadratic programming (QP) problem

$$\begin{aligned} & \text{minimize} \quad \frac{1}{2}x^T Qx + p^T x \\ & \text{subject to} \quad Ax = b, Cx \leq d, x \in X \end{aligned} \tag{1}$$

where $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ is the unknown variable, $Q \in \mathbb{R}^{n \times n}$, $p \in \mathbb{R}^n$, $A \in \mathbb{R}^{r \times n}$, $b \in \mathbb{R}^r$, $C \in \mathbb{R}^{m \times n}$, $d \in \mathbb{R}^m$ are constants, and X is a nonempty box set defined as $X = \{x \in \mathbb{R}^n | l_i \leq x \leq h_i, i = 1, \dots, n\}$ (note that some l_i 's can be $-\infty$ and some h_i 's can be $+\infty$). In addition, Q is symmetric and positive semidefinite. In particular, if $Q = 0$, the problem degenerates to a linear programming (LP) problem.

During the past two decades, many recurrent neural networks have been proposed for solving optimization-related problems because of their analog circuits implementability and intrinsic parallelism which are advantageous for fast computing. In particular, for solving the QP problem (1), there exist many salient models but their real-time computational abilities are often constrained by various deficiencies such as demand for slack variables (for converting inequality constraints into equality constraints) [1, 2, 3, 4, 5, 6], calculation of matrix

inverses [7, 8, 9, 10, 11], inconvenience in determining the convergence conditions [10, 5, 12], and so on. Actually, these networks are suitable for specific applications. Among those capable of solving the QP problem (1) but free of such deficiencies, two simple models refer to

$$\frac{d}{dt} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = -\lambda \begin{pmatrix} x - \mathcal{P}_X((I - Q)x - C^T y + A^T z - p) \\ y - \tilde{y} \\ Ax - b \end{pmatrix} \quad (2)$$

and

$$\frac{d}{dt} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = -\lambda \begin{pmatrix} 2(x - \mathcal{P}_X((I - Q)x - C^T \tilde{y} + A^T(z - Ax + b) - p)) \\ y - \tilde{y} \\ Ax - b \end{pmatrix} \quad (3)$$

where $\tilde{y} = (y + Cx - d)^+$ and $\lambda > 0$, which were proposed in [13] and [14], respectively. The block diagram of the former is depicted in Fig. 1 and that of the latter can be depicted similarly.

The major difference between the performances of (2) and (3) is that the former can solve (1) with positive definite Q , while the latter can solve (1) with positive semi-definite Q though its structure is slightly more complicated (a detailed comparison is made in Section 2).

The major difference between the dynamic equations of the two neural networks is that the latter substitutes $(y + Cx - d)^+$ and $z - Ax + b$ for y and z respectively in the first equation of the former. Clearly, such *substitutions* do not affect the equilibrium set of (2) and consequently the two networks have the same equilibrium set. This observation raises an interesting question: will other combinations of variable substitutions also result in feasible neural networks? Note that multiple variable substitutions are available if we let the right-hand-sides of (2) and (3) be equal to zeros. The answer to this question is positive and a novel model has been figured out in this way very recently [15], which shares the same performance with (3) and the same structural complexity with (2). In this paper we present another model, also resulted from this idea. It will be shown to possess the merits of (2) and (3), and thus can compete with the model in [15].

Throughout the paper, if a is a vector, then $\|a\| = \sqrt{\sum a_i^2}$. \mathfrak{R}_+^n denotes the nonnegative quadrant of \mathfrak{R}^n . Moreover, it is assumed that there exists at least one finite solution to problem (1).

2 Formulation of the Model

As for problem (1), the Karush-Kuhn-Tucker (KKT) conditions can be expressed in the following way [13, 14]: x is an optimum of (1) if and only if there exist $y \in \mathfrak{R}^m$ and $z \in \mathfrak{R}^r$ so that

$$\begin{cases} x = \mathcal{P}_X((I - Q)x - C^T y + A^T z - p) \\ y = (y + Cx - d)^+ \\ Ax - b = 0 \end{cases} \quad (4)$$

where $\mathcal{P}_X(\cdot)$ and $(\cdot)^+$ are two activation functions, whose definitions can be found in any of the references [2, 14, 13, 5, 11, 16, 15, 17]. Obviously, the identical equilibrium set of (2) and (3) coincides with the KKT point set of problem (1).

If we substitute y in the first equation of (2) with $(y + Cx - d)^+$, which is available from the second equation, to constitute \tilde{y} , and then substitute x in the third equation of (2) with \tilde{x} , we arrive at the following dynamic equations with only the scaling factors different:

$$\frac{d}{dt} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = -\lambda \begin{pmatrix} 2(x - \tilde{x}) \\ y - \tilde{y} \\ 2(A\tilde{x} - b) \end{pmatrix} \quad (5)$$

where $\tilde{y} = (y + Cx - d)^+$, $\tilde{x} = \mathcal{P}_X((I - Q)x - C^T\tilde{y} + A^Tz - p)$ and $\lambda > 0$. These equations represent a new network for solving problem (1). The output can be regarded as either x or \tilde{x} because in Section 3 it will be shown that the state equations will always evolve to one of the equilibrium points and at any equilibrium point x is equal to \tilde{x} . It is easy to verify the following results.

Theorem 1. *A point x^* is a solution of (1) if and only if there exist y^* and z^* such that $((x^*)^T, (y^*)^T, (z^*)^T)^T$ is an equilibrium point of (5).*

Then, to verify the viability of the new model, we need to show whether its stability conditions and structural complexity are comparable with existing models especially (2) and (3). In Section 3 it will be shown that the stability results of the proposed model requires only the positive semidefiniteness of Q for solving (1), the same as the neural network (3). In the rest of this section, we compare its structural complexity with those of (2) and (3).

First, we show that the structural complexity of the proposed neural network is the same as (2). One would argue that if the detailed expressions of \tilde{y} and \tilde{x} are substituted into (5), the equations in (5) will get much longer than the corresponding ones in (2); therefore, the structure of the network (5) is much more complicated than that of (2). However, nobody would do that in hardware implementation. Actually, though \tilde{y} appears three times in (5), it does not mean that this term is needed to be implemented three times. Only one block is enough and its output can be connected to three different spots. Likewise, though \tilde{x} appears twice in (5), it requires just one implementation. Based on this idea, on one hand, if we count the numbers of multiplications and additions/subtractions to be performed on the right-hand-sides of (2) and (5), respectively, it will be found that the numbers are the same for the two equations. On the other hand, if we count the numbers of integrators (for realizing integrations), amplifiers (for realizing activation functions $\mathcal{P}_X(\cdot)$ and $(\cdot)^+$), multipliers, summators and interconnections required by hardware implementation of the two networks, respectively, it will be found that all of these numbers are identical for two networks. This result is made clearer in Fig. 1. In the figure, if we change the starting positions of two connections in the neural network (2) to other two positions, and meanwhile change two scaling factors, the figure becomes exactly the diagram of the neural network (5). In this sense, the proposed neural network possesses the same structural complexity as the neural network (2).

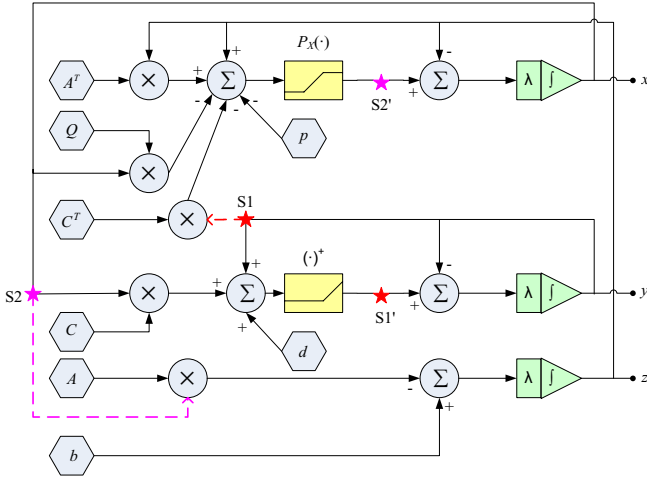


Fig. 1. Block diagrams of the neural networks (2) and (5). If the two dashed lines are connected, the figure depicts the diagram of (2). If the first dashed line (counted top-down, the same convention is adopted in what follows) is not started from point S1, but from point S1', the second line is not started from point S2, but from point S2', while the first and third integrators use 2λ as their scaling factors, the diagram then switches to the network (5).

Let us then examine the complexity of the neural network (3). From its dynamic equations, it is easy to see that no matter how the computing order of different terms is arranged, one more operation of addition is required to incorporate $-Ax + b$ to the term $A^T(z - Ax + b)$ in the top equation of (3), though the result of $Ax - b$ can be obtained directly from the bottom equation. This leads to r more connections and one more summator in its architecture than in the architectures of (2) and (5), which can be perceived in Fig. 1. When the number of equality constraints is large relative to inequality constraints, e.g., in k -WTA applications [8, 15], the inefficiency of the network (3) becomes prominent.

3 Stability Analysis

In this section, we will show that the proposed neural network (5) possesses the same convergence result as the neural network (3), that is, the global convergence property requires the positive semidefiniteness of Q only. In what follows, the equilibrium set of the neural network is denoted by Ω^* . Since it is assumed that (1) has at least one finite solution, according to Theorem 1, there exists at least one finite point in Ω^* . The following lemma follows from [14].

Lemma 1. *The function $\|\tilde{y}\|^2$ is convex and continuously differentiable on \mathbb{R}^{n+m} . In addition, $\nabla\|\tilde{y}\|^2 = 2 \begin{pmatrix} C^T \tilde{y} \\ \tilde{y} \end{pmatrix}$.*

Lemma 2. Consider the following function

$$V(u) = \phi(u) - \phi(u^*) - (u - u^*)^T \nabla \phi(u^*) + \frac{1}{2} \|u - u^*\|^2, \quad (6)$$

where $u = ((x)^T, (y)^T, (z)^T)^T$ and $u^* = ((x^*)^T, (y^*)^T, (z^*)^T)^T \in \Omega^*$ is a finite point and $\phi(u) = x^T Qx/2 + p^T x + \|\tilde{y}\|^2/2$. It has the following properties.

1. $V(u)$ is convex and continuously differentiable on \mathfrak{R}^{n+m+r} .
2. $V(u) \geq \|u - u^*\|^2/2$ for all $u \in \mathfrak{R}^{n+m+r}$.
3. $\nabla V(u)^T F(u) \geq 2\|x - \tilde{x}\|^2 + \|y - \tilde{y}\|^2$ for all $u \in \mathfrak{R}^{n+m+r}$, where

$$F(u) = \begin{pmatrix} 2(x - \tilde{x}) \\ y - \tilde{y} \\ 2(A\tilde{x} - b) \end{pmatrix}.$$

Proof. From Lemma 1, $V(u)$ is continuously differentiable on \mathfrak{R}^{n+m+r} . To prove its convexity, what we only need to show is the convexity of the function $\psi(u) = \phi(u) - \phi(u^*) - (u - u^*)^T \nabla \phi(u^*)$. In view that $\phi(u)$ is convex and $\nabla \psi(u) = \nabla \phi(u) - \nabla \phi(u^*)$, it is easy to validate this fact.

2) This result follows directly from the convexity of $\phi(u)$, which ensures $\psi(u)$ defined above is nonnegative.

3) The gradient of $V(u)$ is given by

$$\nabla V(u) = \begin{pmatrix} (I + Q)(x - x^*) + C^T \tilde{y} - C^T y^* \\ \tilde{y} - 2y^* + y \\ z - z^* \end{pmatrix}.$$

In the following projection inequality (see [18, 13, 14, 16])

$$(\mathcal{P}_\Omega(u) - u)^T (v - \mathcal{P}_\Omega(u)) \leq 0, \forall u \in \mathfrak{R}^n, v \in \Omega,$$

let $\Omega = X$, $u = x - Qx - p - C^T \tilde{y} + A^T z$ and $v = x^*$, then

$$(\tilde{x} - x^*)^T (x - Qx - p - C^T \tilde{y} + A^T z - \tilde{x}) \geq 0.$$

Since u^* satisfies (4), according to the equivalence of the projection equation and variational inequality [18] we know

$$(\tilde{x} - x^*)^T (Qx^* + p + C^T y^* - A^T z^*) \geq 0.$$

Adding the above two equations gives

$$(\tilde{x} - x^*)^T (x - Qx - C^T \tilde{y} + A^T z - \tilde{x} + Qx^* + C^T y^* - A^T z^*) \geq 0.$$

Similarly, we can derive $(\tilde{y} - y^*)^T (y + Cx - d - \tilde{y}) \geq 0$ and $(\tilde{y} - y^*)^T (-Cx^* + d) \geq 0$. Adding them gives

$$(\tilde{y} - y^*)^T (y + Cx - Cx^* - \tilde{y}) \geq 0.$$

Given these equations, we have

$$\begin{aligned}
 \nabla V(u)^T F(u) &= 2\|x - \tilde{x}\|^2 + 2(x - \tilde{x})^T(\tilde{x} - x^* + Qx - Qx^* + C^T\tilde{y} - C^T y^*) \\
 &\quad + \|y - \tilde{y}\|^2 + 2(y - \tilde{y})^T(\tilde{y} - y^*) + 2(A\tilde{x} - b)^T(z - z^*) \\
 &= 2\|x - \tilde{x}\|^2 + \|y - \tilde{y}\|^2 - 2(\tilde{x} - x^*)^T(\tilde{x} - x^* + Qx - Qx^* + C^T\tilde{y} \\
 &\quad - C^T y^*) + 2(x - x^*)^T(\tilde{x} - x^*) + 2(x - x^*)^T(Qx - Qx^*) \\
 &\quad + 2(x - x^*)^T(C^T\tilde{y} - C^T y^*) + 2(\tilde{y} - y^*)^T(y - \tilde{y}) \\
 &\quad + 2(\tilde{x} - x^*)^T(A^T z - A^T z^*) \\
 &= 2\|x - \tilde{x}\|^2 + \|y - \tilde{y}\|^2 + 2(\tilde{x} - x^*)^T(x - \tilde{x} - Qx + Qx^* - C^T\tilde{y} \\
 &\quad + C^T y^* + A^T z - A^T z^*) + 2(x - x^*)^T(Qx - Qx^*) \\
 &\quad + 2(\tilde{y} - y^*)^T(Cx - Cx^* + y - \tilde{y}) \\
 &\geq 2\|x - \tilde{x}\|^2 + \|y - \tilde{y}\|^2 + 2(x - x^*)^T Q(x - x^*).
 \end{aligned}$$

Since Q is positive semidefinite, the conclusion holds.

Lemma 3. *For any initial point $u(t_0) = (x(t_0)^T, y(t_0)^T, z(t_0)^T)^T \in X \times \mathfrak{R}^{m+r}$, the neural network (5) has a unique continuous solution $x(t)$ for all $t \geq t_0$ and $x(t)$ stays in X forever.*

Proof. Taking Lemma 2 into account, the results can be reasoned following similar arguments for proving Theorem 2 in [14], which are omitted here for brevity.

Theorem 2. *For any $u(t_0) \in X \times \mathfrak{R}^{m+r}$ the neural network (5) is stable in the sense of Lyapunov and its trajectory $u(t)$ converges to a point in Ω^* . In particular, if there is only one point in Ω^* , the neural network is asymptotically stable.*

Proof. According to Lemma 3, for any initial point $u(t_0) \in X \times \mathfrak{R}^{m+r}$, the neural network has a unique continuous trajectory $u(t)$ for all $t \geq t_0$. In addition $u(t)$ is bounded for all $t \geq t_0$. According to Lemma 3, $x(t) \in X$ for all $t \geq t_0$. Consider the function defined in (6). It follows from Lemma 2 that

$$\frac{dV(u(t))}{dt} = -\lambda \nabla V(u)^T F(u) \leq -2\lambda \|x - \tilde{x}\|^2 - \lambda \|y - \tilde{y}\|^2 \quad \forall t \geq t_0.$$

Hence, the neural network is stable in the sense of Lyapunov. Clearly, if $du/dt = 0$, then $dV/dt = 0$. According to the LaSalle invariance principle, $u(t)$ converges to the largest invariant set \mathcal{M} in $\{u \in \mathfrak{R}^{n+m+r} | dV(u)/dt = 0\}$. In what follows we show $\mathcal{M} = \Omega^*$. Clearly, any point in Ω^* also belongs to \mathcal{M} . Consider any point $u \in \mathcal{M}$. Since $dV/dt = 0$, then $x = \tilde{x}$ and $y = \tilde{y}$ from the above equation, which implies $dx/dt = -2\lambda(x - \tilde{x}) = 0$. It follows that x is in the steady state (a constant), so is \tilde{x} . Denote $A\tilde{x} - b$ by c where c is a constant. If $c \neq 0$, then $dz/dt = -2\lambda c$ and $z \rightarrow \infty$ when $t \rightarrow +\infty$, which contradicts the boundedness of $u(t)$. Consequently, $c = 0$ and $dz/dt = 0$. It follows that $u \in \Omega^*$, and hence $\mathcal{M} = \Omega^*$.

Since $u(t)$ is bounded over $[t_0, +\infty)$, there exists a convergent subsequence $t_0 < \dots < t_n < t_{n+1} < \dots$ such that $\lim_{k \rightarrow +\infty} u(t_k) = \hat{u}$, where $\hat{u} \in \Omega^*$. Define another Lyapunov function $\hat{V}(u)$ the same as $V(u)$ in (6) except that u^* in $V(u)$ is replaced by \hat{u} . It is easy to see that $\hat{V}(u)$ decreases along the trajectory of (5) and satisfies $\hat{V}(\hat{u}) = 0$. Therefore, for any $\varepsilon > 0$, there exists $q > 0$ such that, for all $t \geq t_q$,

$$\|u(t) - \hat{u}\|^2/2 \leq \hat{V}(u(t)) \leq \hat{V}(u(t_q)) < \varepsilon,$$

that is, $\lim_{t \rightarrow +\infty} u(t) = \hat{u}$.

In particular, if Ω^* contains a unique point, from the above analysis, the neural network is asymptotically stable. The proof is completed.

Therefore, the proposed model is as excellent as (3) in terms of performance.

4 Illustrative Examples

In this section, we will discuss two examples to demonstrate the performance of the proposed neural network. The simulations were conducted in MATLAB.

Example 1. Consider the following problem discussed in [14]:

$$\begin{aligned} & \text{minimize} && (x_1 + 3x_2 + x_3)^2 + 4(x_1 - x_2)^2 \\ & \text{subject to} && x_1^3 - 6x_2 - 4x_3 + 3 \leq 0, \\ & && 1 - x_1 - x_2 - x_3 = 0, x \geq 0. \end{aligned}$$

This problem has a unique solution $x^* = (0, 0, 1)^T$, and the associated Lagrange multipliers are $y^* = 0, z^* = -2$. It is easy to verify that the problem is convex on $X = \mathbb{R}_+^3$. According to Theorem 2, the neural network (5) should converge to u^* with any initial point $u(0) \in \mathbb{R}_+^3 \times \mathbb{R}^2$ and be asymptotically stable at u^* . All simulations verified this fact. In addition, it was interesting to note that even with $x(0) \notin X$, the neural network still always converged to u^* . Fig. 2 shows such an example. This phenomenon indicates that the stability result in Theorem 2 leaves space for improving.

Example 2. To illustrate the performance of the proposed neural network for linear programming, we now solve a classical transportation problem [19]. Suppose that M suppliers, each with a given amount of goods p_i , are required to supply N consumers, each with a given limited capacity q_j . For each supplier-consumer pair, the cost of transporting a single unit of goods is given as c_{ij} . The transportation problem is then to find a least-expensive flow of goods from the suppliers to the consumers that satisfies the consumers' demand. Formally, the problem can be described as follows

$$\begin{aligned} & \text{minimize} && \sum_{i=1}^M \sum_{j=1}^N c_{ij} x_{ij} \\ & \text{subject to} && \sum_{j=1}^N x_{ij} \leq p_i, \quad \forall 1 \leq i \leq M \end{aligned}$$

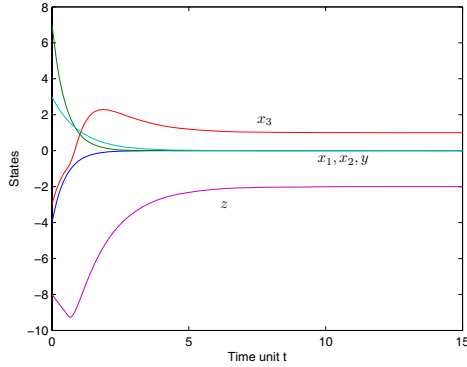


Fig. 2. State trajectory of the neural network (5) with $\lambda = 1$, $u(0) = (-4, 7, -3, 3, -8)^T$ in Example 1

$$\sum_{i=1}^M x_{ij} \leq q_j, \quad \forall 1 \leq j \leq N$$

$$\sum_{i=1}^M \sum_{j=1}^N c_{ij} x_{ij} = \min \left(\sum_{i=1}^M p_i, \sum_{j=1}^N q_j \right)$$

$$x_{ij} \geq 0, \quad \forall 1 \leq i \leq M, 1 \leq j \leq N.$$

The problem can be rewritten in the standard form of LP problem (11) with $Q = 0$, and the neural network (5) can be used to solve the problem. For illustration purpose, let us consider a problem with $M = 3$ and $N = 4$. The parameters are $p = (10, 16, 18)^T$, $q = (13, 5, 15, 10)^T$ and

$$c = \begin{pmatrix} 0.1 & 0.2 & 0.1 & 0.5 \\ 0.5 & 0.1 & 1.0 & 0.8 \\ 1.0 & 0.1 & 0.4 & 0.1 \end{pmatrix}.$$

The unique solution is

$$x^* = \begin{pmatrix} 3 & 0 & 7 & 0 \\ 10 & 5 & 0 & 0 \\ 0 & 0 & 8 & 10 \end{pmatrix}.$$

The neural network (5) ($Q = 0$) was simulated to solve the problem and obtained the same solution. Fig. 3 depicts the output trajectory $x(t)$ in one simulation with $\lambda = 1$ and a random initial point, which converge to x^* .

5 Concluding Remarks

In the paper a new recurrent neural network was presented for quadratic and linear programming. This model shares much similarity with two classical models but combines their merits, that is, simple structure and good performance.

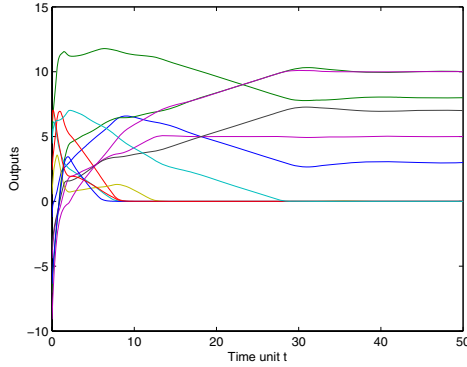


Fig. 3. Output trajectory of the neural network (5) with $\lambda = 1$ and a random initial point $u(0)$ in Example 2

The contribution of this invention is twofold. On one hand, it enriches the family of recurrent neural networks for solving optimization problems, and therefore offers flexibility for circuits practitioners in neural circuits design. On the other hand, its design idea, actually, obtaining new models from existing ones by variable substitutions, may shed light to the development of more powerful models. Up to now, two new models have been devised using this idea (the other is presented in [15]). An interesting question arises: how many models on earth are there in this category besides the known four? We believe that this open question will stimulate many further investigations.

Acknowledgements

The work was supported by the National Natural Science Foundation of China Grants 60805023, 60621062 and 60605003, National Key Foundation R&D Project Grants 2003CB 317007, 2004CB318108 and 2007CB311003, China Postdoctoral Science Foundation Grants 20080430032 and 200801072, and Basic Research Foundation of Tsinghua National Laboratory for Information Science and Technology (TNList).

References

1. Xia, Y.: A New Neural Network for Solving Linear and Quadratic Programming Problems. *IEEE Trans. Neural Netw.* 7, 1544–1547 (1996)
2. Tao, Q., Cao, J., Sun, D.: A Simple and High Performance Neural Network for Quadratic Programming Problems. *Applied Mathematics and Computation* 124, 251–260 (2001)
3. Gao, X., Liao, L.: A Neural Network for Monotone Variational Inequalities with Linear Constraints. *Physics Letters A* 307, 118–128 (2003)

4. Ghasabi-Oskoei, H., Mahdavi-Amiri, N.: An Efficient Simplified Neural Network for Solving Linear and Quadratic Programming Problems. *Applied Mathematics and Computation* 175, 452–464 (2006)
5. Yang, Y., Cao, J.: Solving Quadratic Programming Problems by Delayed Projection Neural Network. *IEEE Trans. Neural Netw.* 17, 1630–1634 (2006)
6. Barbarosou, M.P., Maratos, N.G.: A Nonfeasible Gradient Projection Recurrent Neural Network for Equality-Constrained Optimization Problems. *IEEE Trans. Neural Netw.* 19, 1665–1677 (2008)
7. Zhang, Y., Wang, J.: A Dual Neural Network for Convex Quadratic Programming Subject to Linear Equality and Inequality Constraints. *Physics Letters A* 298, 271–278 (2002)
8. Liu, S., Wang, J.: A Simplified Dual Neural Network for Quadratic Programming with Its KWTA Application. *IEEE Trans. Neural Netw.* 17, 1500–1510 (2006)
9. Hu, X., Wang, J.: Solving Generally Constrained Generalized Linear Variational Inequalities Using the General Projection Neural Networks. *IEEE Trans. Neural Netw.* 18, 1697–1708 (2007)
10. Liu, Q., Wang, J.: A One-Layer Recurrent Neural Network with a Discontinuous Hard-Limiting Activation Function for Quadratic Programming. *IEEE Trans. Neural Netw.* 19, 558–570 (2008)
11. Hu, X., Wang, J.: Design of General Projection Neural Networks for Solving Monotone Linear Variational Inequalities and Linear and Quadratic Optimization Problems. *IEEE Trans. Syst. Man, Cybern. B* 37, 1414–1421 (2007)
12. Forti, M., Nistri, P., Quincampoix, M.: Generalized Neural Network for Nonsmooth Nonlinear Programming Problems. *IEEE Trans. Circuits Syst. I* 51, 1741–1754 (2004)
13. Xia, Y.: An Extended Projection Neural Network for Constrained Optimization. *Neural Computation* 16, 863–883 (2004)
14. Gao, X.: A Novel Neural Network for Nonlinear Convex Programming. *IEEE Trans. Neural Netw.* 15, 613–621 (2004)
15. Hu, X., Zhang, B.: A New Recurrent Neural Network for Solving Convex Quadratic Programming Problems with an Application to the k -Winners-Take-All Problem. *IEEE Trans. Neural Netw.* (accepted)
16. Hu, X., Wang, J.: An Improved Dual Neural Network for Solving a Class of Quadratic Programming Problems and Its K -Winners-Take-All Application. *IEEE Trans. Neural Netw.* 19, 2022–2031 (2008)
17. Wang, Z., Perterson, B.S.: Constrained Least Absolute Deviation Neural Networks. *IEEE Trans. Neural Netw.* 19, 273–283 (2008)
18. Kinderlehrer, D., Stampcchia, G.: *An Introduction to Variational Inequalities and Their Applications*. Academic, New York (1980)
19. Hitchcock, F.L.: The Distribution of a Product from Several Sources to Numerous Localities. *J. Math. Phys.* 20, 224–230 (1941)

A Particle Swarm Optimization Algorithm Based on Genetic Selection Strategy

Qin Tang^{1,2,*}, Jianyou Zeng³, Hui Li⁴,
Changhe Li⁵, and Yong Liu⁶

¹ Department of Control Science and Engineering,
Huazhong University of Science and Technology, Wuhan 430074, China

² School of Mathematics and Physics,
China University of Geosciences, Wuhan 430074, China

³ School of Art and Communication,
China University of Geosciences, Wuhan 430074, China

⁴ School of Computer, China University of Geosciences, Wuhan 430074, China

⁵ Department of Computer Science, University of Leicester
Leicester LE1 7RH, UK

⁶ Faculty of Mechanical and Electronic Information,
China University of Geosciences, Wuhan 430074, China
tq7171@gmail.com, jianyou@cug.edu.cn, huili@vip.sina.com,
chli@cug.edu.cn, adamlau430@yahoo.com.cn

Abstract. The standard particle swarm optimization algorithm (simply called PSO) has many advantages such as rapid convergence. However, a major disadvantage confronting the PSO algorithm is that they often converge to some local optimization. In order to avoid the occurrence of premature convergence and local optimization of the PSO algorithm, a particle swarm optimization algorithm based on genetic selection strategy, simply called GSS-PSO, is singled out in this paper. GSS-PSO not only retains the rapid convergence charactering of the standard PSO algorithms, but also scales up their global search ability. At last, we experimentally tested the efficiency of our new GSS-PSO algorithm using eight classical functions. The experimental results show that our new GSS-PSO algorithm is generally better than the PSO algorithm.

Keywords: PSO, Convergence, Function optimization, GSS.

1 Introduction

PSO [1] is a random global search method which was proposed by American socio-psychologists Eberhart and Kennedy, who were inspired by the relevant

* Qin Tang received the B.Sc. in Resource Prospecting Engineering in 1992 and M.Sc. in Scientific and Technology History in 2002 from China University of Geosciences, China. Currently, she is pursuing the PhD Degree in Department of Control Science and Engineering Huazhong University of Science and Technology, China. Her research interest includes multiple objective particle swarm optimization and neural networks.

researches in the field of swarm intelligence and formed an evolutionary computation technique. In PSO, the population refers to a set of potential solutions. The initial species are formed randomly and uniform distributed, and then an iterative search is conducted by the population in the solution space through the simulation of social and cognitive process. In the process of iterative search, a global information exchange is conducted among all the particles. By the information exchange, each particle can share the current searching results of other particles. In this way, the particles can modify the knowledge of the searching space to achieve the goal of common evolution of the whole population. Compared with GAs, which is also operated on the basis of population, PSO does not depend on the genetic operators such as selection operator, crossover operator, mutation operator to manipulate each particle [2], but starts with a group of random solutions to get the optimal solution through information exchange and iterative search. PSO attracts the attention of the researchers because of its advantages such as simple conception, easy realization, dependent just on function value without the gradient information of the target function, rapid convergence, and is applied to extensive areas, including multi-objective optimization problems, minimum-maximum problems, integer programming and many application problems in the actual projects. A number of researches and experiments prove that PSO is an efficient optimization technique [3,4]. However, the standard PSO may reduce the population diversity and limit the solution to certain local optimization points, failing to obtain the global optimization due to the individual conformity, which occurs in the process of information exchange between the individual and population during iterative search. This phenomenon is generally defined as premature convergence, which is a severe problem of evolutionary algorithm techniques. The main reason is that the high pressure of selection and rapid convergence cause the loss of the ability of the particles to exploit new areas, consequently the algorithm is limited to local optimization [5]. This thesis offers some modifications according to the defect of local optimization of the standard PSO by involving genetic selection strategy to enhance the global searching ability of the particles. The result of experiments demonstrates the advantage of GSS-PSO over the standard PSO.

2 The Basic PSO Algorithm

Kennedy and Elberhart modified the simulation system of bird's community BOID [6], which was proposed by Reynolds, and added a specific point which was defined as food. The birds search for their food according to the foraging behavior of the surrounding birds. They intended to imitate the way birds searched for food by this model. Surprisingly, the result of the experiment indicated tremendous optimizing ability of this simulating model, especially in multi-dimensioned space. In this imitating system, the term "particle" is used to represent each individual particle. Every particle has a fitness value which is determined by the target function and a speed which governs their flying direction and distance. All the particles search in the solution space according

to their own optimal positions and the position of the best particle among the whole population. PSO is initiated as a swarm of random solutions and gets the optimal solution by iterative search. During each iterative search, the particles update themselves by tracing two extreme values: one is the optimal positions of the particles themselves, which is defined as *pbest*; the other is the current optimal position of the whole population, which is called *gbest*. The particles determine their next movements according to the experience of their own and other particles.

PSO can be described as follows: supposing there is an N-dimensional searching space and m particles in the population. The i th particle can be described as a N-dimensional vector, with $X_i = (x_{i1}, x_{i2}, \dots, x_{iN})$ $i = 1, 2, \dots, m$, i.e. the position of the i th particle in N-dimensional space is X_i , the best position it occupies if $P_i = (p_{i1}, p_{i2}, p_{iN})$ $i = 1, 2, m$. Each position of the particle represents a potential solution to the existing problem. The fitness value is obtained by putting the position value into the target function to judge the quality of the particle. The current optimal position of the whole population is described as $P_g = (p_{g1}, p_{g2}, p_{gN})$, in which g is the index of the position of the best particle.

As to the particle id , its updating formulations of speed and position are as follows:

$$V'_{id} = \omega V_{id} + \eta_1 \text{rand}() (P_{idb} - X_{id}) + \eta_2 \text{rand}() (P_{gdb} - X_{id}) \quad (1)$$

$$X'_{id} = X_{id} + V'_{id} \quad (2)$$

where ω is the inertia weight, which is the proportional factor relating to the previous speed, $0 < \omega < 1$; η_1 and η_2 are constants, called accelerate factors, and generally speaking $\eta_1 = \eta_2 = 2$; $\text{rand}()$ is a random number; X_{id} represents the position of particle id , while V_{id} means the speed of id ; P_{idb} and P_{gdb} represent the current best position particle id finds and the position of the best particle among the population respectively. The speed of each particle is constantly adjusted during the iterative search. Accelerate factor $\eta_1 \text{rand}()$ and $\eta_2 \text{rand}()$ are two random factors, and the random number on each dimension is formed respectively to balance the weight which pulls the particle to P_i and P_g . Controlling these two accelerate factors can change the mutual effects between the current best position of a certain particle and the current best position of its adjacent particles. A bigger value will cause the particle to fly more vibrantly toward or even over the target area.

The first part in (1) is the previous speed of the particle, and this speed enables the particle with an expanding tendency in the searching space to give the algorithm a global searching ability; the second is the cognitive component which demonstrates the process of the particle taking in the information of its own experience; the third is the social component, showing the process of learning the experience from other particles and indicating the information sharing and social cooperation among the particles. All these three parts determine the spatial searching ability of the particle together: the first part balances the global and local searching capacity, and the second part empowers the particle with a strong global searching ability to avoid local minimum, while the third part shows the

share of information among all the particles. Only under the co-effect of these three parts can the particles achieve the goal of best position efficiency.

One outstanding features of PSO is the rapid convergence of the population, and Clerc [7] also provided evidence of its efficient convergence. However, it is this advantage that causes its fatal defect: with the increase of iterative searches, the speed of the particles will be reduced gradually and finally the value tends to be zero, and at this moment the whole population is converged at a certain point in the space. But if the best particle does not find the global optimal position, the population will turn out to be local optimization, and the possibility of getting rid of local optimization is rather slim. This situation will happen more frequently especially to multimodal function. The comparative experiment of multimodal function f_8 in this thesis also proves the result, as shown in Table 1 and Fig.1. In order to overcome this defect of PSO, the GSS-PSO algorithm is proposed in this thesis.

3 The GSS-PSO Algorithm

The standard PSO is efficient in convergence, and the adjustment of cognitive and social components enables the particles to search centered round the two optimal values P_{gd} and P_{id} . Once the best individual of the population is limited to local optimization, according to the exchange formulation of speed and position, the information sharing mechanism of PSO will attract the other particles to move forward the local optimal position during the following searching process, finally the whole population will converge in this position. When the population is limited to local optimization, according to the speed updating (1), the cognitive and social components will be zero, and with the increase of iterative search, the speed of the particle will tend to be zero because of the limitation $0 < \omega < 1$. Under this circumstance, the possibility of the population to overcome local optimization will be too slim to get the global optimization solution. In order to avoid local optimization, GSS-PSO adopts a new information sharing mechanism: as is known to all, in the process of solution searching, none of the particles knows the position of the optimal solution. However, the best position each particle ever occupied and the best position the particle swarm ever occupied can still be recorded; moreover, the worst position the particle and the particle swarm ever occupied also can be recorded. In this way, the particles can be controlled not to move toward the worst position itself (p_{worst}) and the global population (g_{worst}) occupied, therefore the global searching space is enlarged. This method avoids the problem of an early local optimization and enhances the possibility to get the global optimization solution during the search in solution space.

The updating speed and position formulation of the particle based on the new strategy is as follows:

$$V'_{id} = \omega V_{id} + \eta_1 rand()(X_{id} - P_{idw}) + \eta_2 rand()(X_{id} - P_{gdw}) \quad (3)$$

$$X'_{id} = X_{id} + V'_{id} \quad (4)$$

Where, P_{idw} and P_{gdw} represent the worst positions of particle id and that of the global population respectively.

The next flying direction of each particle in the standard particle group is almost certain, which means that the particle can only move toward the ever best position of its own and of the whole population. The particle is easy to be limited to local optimization by the above mentioned behavior analysis. In order to further reduce the possibility of local optimization, the genetic selection strategy is adopted in GSS-PSO: supposing the number of the particle is m , and there will be m sub-generations according to PSO. Additionally, there will still be m sub-generations by (3) and (4). Therefore, there will be $2m$ sub-generations, which mean m pairs of particles. The fitness values of the two particles in each pair are compared and the particle with a smaller fitness value will enter the next generation. Experiments prove that this new strategy extremely reduces the possibility of local optimization when searching the solutions of some functions.

The procedure of GSS-PSO is as follows:

Step 1: randomly initializing the position and speed of the particles.

Step 2: evaluating the fitness value of each particle.

Step 3: to each particle, updating the best position P_{idb} of particle id, if the fitness value is smaller than the previous best fitness value P_{idb} ; or updating P_{idw} if the fitness value is bigger than that of the previous worst position P_{idw} .

Step 4: to each particle, updating the best position P_{gdb} of particle id, if the fitness value is smaller than the previous best fitness value P_{gdb} ; or updating P_{gdw} if the fitness value is bigger than that of the previous worst position P_{gdw} .

Step 5: to each particle, A new particle t is generated according to (1), and (2). Another new particle t' is generated according to (3) and (4). t and t' are compared and the better one will go into the next generation.

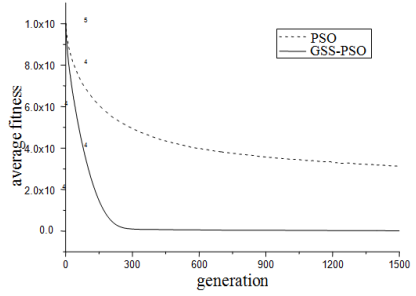
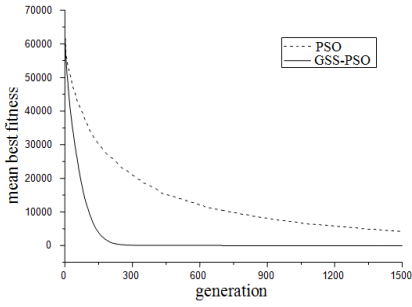
Step 6: a new generation of particles is generated according to the above general selection strategy.

Step 7: stop the procedure when the shutdown strategy is met; otherwise, turn back to step 3.

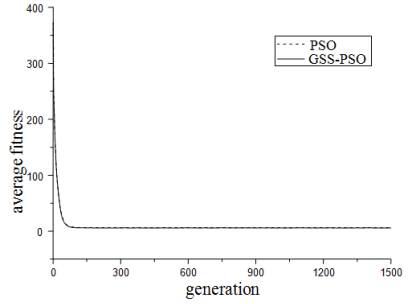
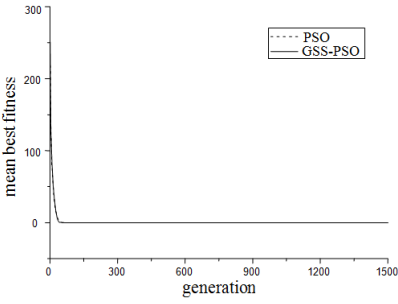
4 Experiments

In this section, we conduct our experiments to compare the efficiency of the PSO algorithm and our new GSS-PSO algorithm. In our experiments, eight benchmark functions are used. The detailed information of the testing functions is shown in Table 1, where $f_1 - f_3$ are single-modal functions, with f_2 a looping function and a noise function U whose uniform distribution scope is (0,1) is added into f_3 . $f_4 - f_8$ are multimodal functions, with n representing the number of dimensions of the functions, S the value range of the variants and f_{min} the minimum value of the functions.

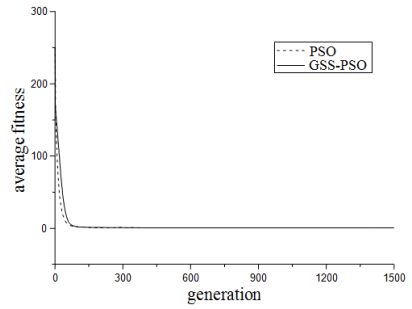
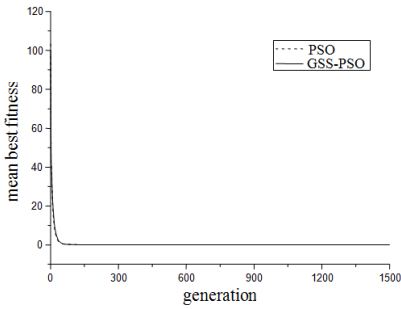
Besides, the parameter settings of the experiments are the same. Based on the researches of van den Bergh[8], parameter design is: $\eta_1 = \eta_2 = 1.496180$, $\omega = 0.729844$. Each testing function is operated 50 times and the statistical results of the experiments are demonstrated in Fig. 1 and Table 2.



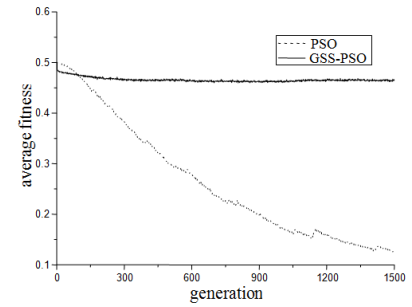
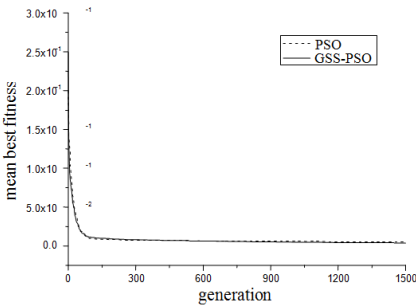
f_1



f_2



f_3



f_4

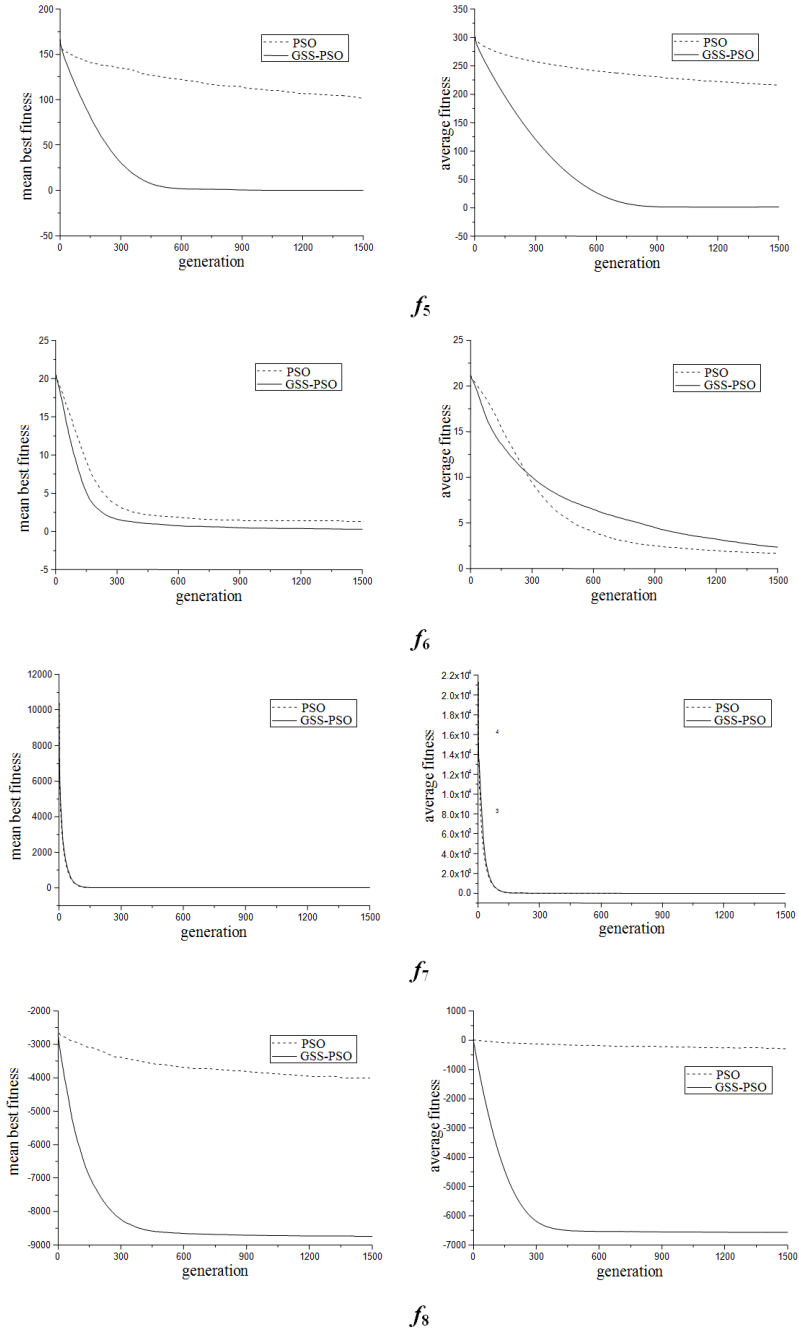


Fig. 1. A comparative experiment of GSS-PSO and PSO (ordinate in the left column represents the average fitness value of the best particles, while ordinate in the right column shows the mean fitness value of the population)

Table 1. Testing functions

Testing functions	n	S	f_{min}
$f_1 = \sum_{i=1}^n x_i^2$	30	(-100,100)	0
$f_2 = 6 \sum_{i=1}^5 x_{i \downarrow}$	30	(-5.12,5.12)	0
$f_3 = \sum_{i=1}^n i = 1^n i x_i^4 + U(0, 1)$	30	(-1.28,1.28)	0
$f_4 = \frac{\sin^2 \sqrt{x^2+y^2+0.5}}{(1.0+0.001(x^2+y^2))^2} + 0.5$	2	(-100,100)	0
$f_5 = \frac{1}{400} \sum_{i=1}^n (x_i - 100)^2 - \prod_{i=1}^n \cos(\frac{x_i-100}{\sqrt{i}}) + 1$	30	(-300,300)	0
$f_6 = -20 \exp(-0.2 \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}) - \exp(\frac{1}{n} \sum_{i=1}^n \cos(2\pi x_i)) + 20 + e$	30	(-32,32)	0
$f_7 = \sum_{i=1}^n 100((x_{i+1} - x_i^2)^2 + (x_i - 1)^2)$	30	(-2.048,2.048)	0
$f_8 = \sum_{i=1}^n -x_i \sin(\sqrt{ x_i })$	30	(-500,500)	-12569.5

Table 2. A comparative experiment of GSS-PSO and PSO

Functions	Algorithms	The best value	Average value	Standard Deviation	The worst value	t-test ^a GSSPSO-PSO
f_1	PSO	1495.71	4224.775	201.038	7032.89	-20.2
	GSS-PSO	4.13731E-29	4.46015E-26	2.28015E-26	1.0882E-24	
f_2	PSO	0	0	0	0	0
	GSS-PSO	0	0	0	0	
f_3	PSO	0.00177094	0.004630085	0.000210055	0.00833963	14.09
	GSS-PSO	0.00193565	0.004941903	0.000259903	0.0103595	
f_4	PSO	0	0.005246591	0.000684817	0.00971591	-1.85
	GSS-PSO	1.41681E-07	0.003697618	0.000509981	0.00971591	
f_5	PSO	72.5069	101.410452	1.52289	123.954	-65.90
	GSS-PSO	2.18559E-12	0.010377	0.00177512	8.63194E-25	
f_6	PSO	-3.19744E-14	1.306447538	0.148418	4.4229	-6.75
	GSS-PSO	-3.19744E-14	1.306447538	0.148418	4.4229	
f_7	PSO	1.84889E-28	4.59643E-26	1.83991E-26	8.63194E-25	1.23
	GSS-PSO	2.55147E-28	3.46992E-25	2.41678E-25	1.20401E-23	
f_8	PSO	-5038.62	-4005.02	54.0123	-3233.13	-66.49
	GSS-PSO	-9535.19	-8741.27	45.1661	-8203.56	

^a The t value is calculated under the condition of a 49 degree of freedom and a 0.05 significant level α .

Fig. 1 and Table 2 indicate that the efficiency of GSS-PSO and PSO are almost the same when solving functions f_2, f_3, f_4 and f_7 , while in f_1, f_5, f_6 and f_8 , the solution of GSS is much better than that of PSO and the same happens to the speed of convergence because of the new adopted information sharing mechanism in GSS-PSO.

Table 2 indicates that GSS-PSO and PSO can both get the global optimal solutions in f_2 ; they get the approximate global optimal solutions in f_3, f_4 , and f_7 ; while GSS-PSO gets better solutions than PSO does in f_1, f_5, f_6, f_8 , and it gets approximate global optimal solution with an exception of f_8 . Fig. 1 also indicates that GSS-PSO is faster in convergence than PSO and the average best

fitness value of GSS-PSO is also much better than that of PSO. On early operating stage, GSS-PSO finds an approximate optimal point with a faster speed and then moves towards the global optimal point. Compared with GSS-PSO, PSO is slower in convergence, due to the over convergence of PSO, which causes the loss of population diversity and limits the population to a local optimization. On the aspect of the average fitness value of the whole population, the convergence of GSS-PSO and PSO is consistent with the average optimal fitness value.

To sum up, when solving some single-modal and multi-modal functions, the information sharing mechanism in GSS-PSO is effective in ensuring the population diversity, and ensures a more advanced global searching ability than PSO.

5 Conclusion

In this paper, we single out a particle swarm optimization algorithm based on genetic selection strategy (simply called GSS-PSO). Our experimental results show that our new GSS-PSO algorithm is generally better than the PSO algorithm in terms of accuracy and rapidity.

In our new GSS-PSO algorithm, two improved ideas are used: 1) A new information sharing mechanism is adopted to record the current worst positions each particle ever occupied and the particle group ever occupied, in this way, the particles can be controlled to fly toward its own ever best position and the ever best position of the whole population to avoid moving toward the worst position of the particle itself ever occupied and the group did. Therefore, the global searching capacity of the particles is enhanced to avoid an early local optimization and increase the possibility to search for a global optimal solution. 2) The genetic selection strategy is adopted to select the next generation of particle species by randomly competing and grading, in order to increase population diversity and reduce the possibility of being limited to local optimization. GSS-PSO keeps the advantages of PSO, such as a simple computation, a speedy convergence; meanwhile, it expands the searching space with a computation method of relatively low complexity.

Acknowledgement

The authors would like to thank the two anonymous reviewers for their valuable comments on earlier versions of this paper.

References

1. Kennedy, J., Eberhart, R.C.: Particle Swarm Optimization. In: Proceedings of IEEE International Conference on Neural Networks, Piscataway, New Jersey, pp. 1942–1948 (1995)
2. Saravanan, N., Waagen, D.E., Eiben, A.E.: Genetic Algorithms and Particle Swarm Optimization. In: Porto, V.W., Waagen, D. (eds.) EP 1998. LNCS, vol. 1447, pp. 611–616. Springer, Heidelberg (1998)

3. Clerc, M., Kennedy, J.: The Particle Swarm - Explosion, Stability, and Convergence in a Multidimensional Complex Space. *IEEE Transactions on Evolutionary Computation* 6(1), 58–73 (2002)
4. Coello Coello, C.A., Lechuga, M.S.: MOPSO: A Proposal for Multiple Objective Particle Swarm Optimization. In: *Proceedings of the IEEE Congress on Evolutionary Computation (CEC 2002)*, Honolulu, Hawaii, pp. 1051–1056 (2002)
5. Fang, W., Kaiyou, L., Yuhui, Q.: A Diversity Strategy for Particle Swarm Optimization. *Computer Science* 33(1), 213–215 (2006)
6. Reynolds, C.W.: Flocks, Herds and Schools: A Distributed Behavioral Model. *Computer Graphics* 22(4), 25–34 (1987)
7. Clerc, M., Kennedy, J.: The Particle Swarm: Explosion, Stability and Convergence in a Multi-Dimensional Complex Space. *IEEE Transactions on Evolutionary Computation* 6(1), 58–73 (2002)
8. Van Den Bergh, F.: An Analysis of Particle Swarm Optimizers. Ph.D's Dissertation Department of Computer Science, University of Pretoria, South Africa (2002)

Structure Optimization Algorithm for Radial Basis Probabilistic Neural Networks Based on the Moving Median Center Hyperspheres Algorithm

Ji-Xiang Du^{1,2} and Chuan-Min Zhai¹

¹ Computer Science & Engineering Dept., Huaqiao University,
Quanzhou 362021, China

² Department of Automation, University of Science and Technology of China,
Hefei, 230026, China
{jxdu77, cmzhai}@gmail.com

Abstract. In this paper, a novel structure optimization algorithm for radial basis probabilistic neural networks (RBPNN) is proposed. Firstly, a moving median center hyperspheres (MMCH) algorithm is proposed to heuristically select the initial hidden layer centers of the RBPNN, and then a hybrid optimization algorithm is adopted to further prune the initial structure of the RBPNN. Finally, the effectiveness and efficiency of our proposed algorithm are evaluated through a plant species identification problem and a palmprint recognition task.

Keywords: Radial basis probabilistic neural networks, Moving median center hyperspheres algorithm, Structure optimization.

1 Introduction

The radial basis probabilistic neural networks (RBPNN) model, as shown in Fig. 1, integrates the advantages of radial basis function neural networks (RBFNN) and probabilistic neural networks (PNN), and avoids or reduces the disadvantages of the RBFNN and the PNN [1]. The construction of a RBPNN involves four different layers: one input layer, two hidden layers and one output layer. The first hidden layer is a nonlinear processing layer, which generally consists of hidden centers selected from a training samples set. The second hidden layer selectively sums the outputs of the first hidden layer, which generally has the same size as the output layer for a labeled pattern classification problem. In general, the weights between the first and the second hidden layer are set as fixed values (1 or 0) and do not require learning. Generally, the first hidden layer is tightly interrelated to the performance of the RBPNN.

Just as for the RBFNN, in the first hidden layer of the RBPNN, the hidden centers number and locations as well as the controlling parameters of the kernel function are quite important indices. Too many hidden centers will lead to very lengthy training and testing time, and poor generalization capability, while, too few hidden centers can lead to quite great convergent error. In addition, the selected hidden centers will require especial controlling parameters in order to realize the entire overlay of training samples in space. The tightly correlative characteristic between the hidden centers and

controlling parameters shows that while investigating the structure optimization for the RBPNN, the hidden centers (including number and locations) and the controlling parameters must be simultaneously considered. Therefore, this paper will discuss how to optimize the full structure of the RBPNN to improve the classification performance and generalization capability of the networks.

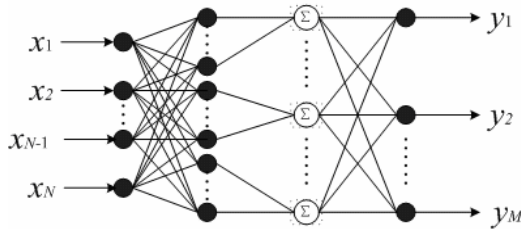


Fig. 1. The topology scheme of the RBPNN

This paper is organized as follows: in Section 2, the full structure optimization algorithm for RBPNN is discussed in details. The experimental results are presented in Section 3, and Section 4 concludes the whole paper and gives related conclusions.

2 Optimizing RBPNN Using Moving Median Center Hyperspheres Algorithm

2.1 Moving Median Center Hyperspheres (MMCH) Algorithm

The fundamental idea of the MMCH method is that each class of patterns can be regard as a series of “hyper spheres”, while in conventional approaches these patterns from one class are all treated as a set of “points”. The first step of this method is to compute the multidimensional median of the points of the considered class, and set the initial center as the closest point from that class to that median. Then find the maximum radius that can encompass the points of the class. Through certain iterations, we remove the center of the hypersphere around in a way that would enlarge the hypersphere and have it encompass as many points as possible. This is performed by having the center “hop” from one data point to a neighboring point. Once we find the largest possible hypersphere, the points inside this hypersphere are removed, and the whole procedure is repeated for the remaining points of the class. We continue until all points of that class are covered by some hyperspheres. At that point, we tackle the points of the next class in a similar manner.

So, the above hypersphere method is used to compress the training data. In other words, the data points of different classes should be covered with a set of compact hyperspheres using the MMC method mentioned above. Thus, these hypersphere centers will serves as the initial hidden layer centers of RBPNNs.

Here, we take one class for example to summarize the whole iterating procedure of the MMC hypersphere classifier as follows; an illustration of the algorithm is shown in Fig.2.

- Step 1. Put all the training data points into a set (we can consider a training data as a point in the hyperspace)
- Step 2. Select the closest point to the median of points in the set as the initial center of the hypersphere.
- Step 3. Find the nearest point of the initial center from all other classes, and denote the distance as d_1 .
- Step 4. Repeat the following steps to enlarge the hypersphere:
 - Step 4.1 Find the farthest point of the same class inside the hypersphere of the radius d_1 to the center. Let d_2 denote the distance from the center to that farthest point.
 - Step 4.2 Set the radius of the hypersphere as $(d_1+d_2)/2$.
 - Step 4.3 Select the point in the most negative direction of the center to the nearest point of the other classes among the nearest k points in this class. If no point can be selected, stop the loop and go to Step 5.
- Step 5. Remove those points encompassed by the hypersphere from the set. If the set is still not empty go to Step 2, else continue.
- Step 6. Remove the redundant hyperspheres that are totally enclosed by larger hyperspheres of the same class.

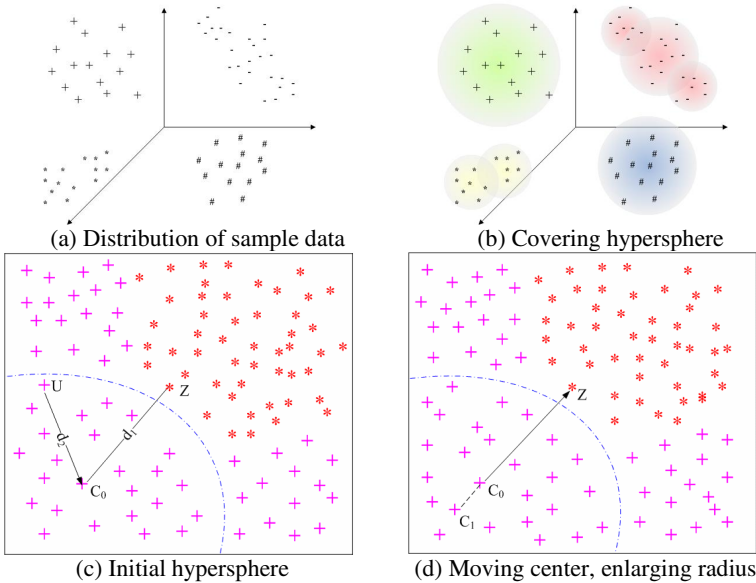


Fig. 2. Illustration of the MMCH algorithm

2.2 Selecting the Hidden Centers by the Hybrid Optimization Algorithm

Assume that \mathbf{Y}_d , \mathbf{H} , \mathbf{W} and $J(\mathbf{W})$, respectively, denote the desired signal matrix, the output matrix of the second hidden layer, the weight matrix between the second

hidden layer and the output layer and the cost function of the RBPNN. In the form of Euclidean norm, the cost function of the RBPNN can be given by

$$\begin{aligned} \mathbf{Y}_w &= \mathbf{Y}_d - \mathbf{H}\mathbf{W} \\ J(\mathbf{W}) &= \|\mathbf{Y}_w\|_2^2 = \sum_{i=1}^d y_{wi}^2 \end{aligned} \quad (1)$$

By conducting the orthogonal decomposition operation, we have:

$$\begin{aligned} \mathbf{H} &= \mathbf{Q}[\mathbf{R}, 0]^T \\ \mathbf{Q}^T \mathbf{Y}_d &= [\tilde{\mathbf{Y}} \quad \tilde{\mathbf{Y}}] \end{aligned} \quad (2)$$

$$J(\mathbf{W}) = \left\| \mathbf{Q} \left[\begin{array}{c} \tilde{\mathbf{Y}} \\ \tilde{\mathbf{Y}} \end{array} \right] - \left[\begin{array}{c} \mathbf{R} \\ 0 \end{array} \right] \mathbf{W} \right\|_2^2 = \|\tilde{\mathbf{Y}} - \mathbf{R}\mathbf{W}\|_2^2 + \|\tilde{\mathbf{Y}}\|_2^2 \quad (3)$$

where $\|\tilde{\mathbf{Y}}\|_2^2$ is the residual error (RE) of $J(\mathbf{W})$, and it is also written as:

$$E_R = \|\mathbf{Y}_d - \mathbf{H}\mathbf{W}\|_2^2 = \|\tilde{\mathbf{Y}}\|_2^2 \quad (4)$$

For pattern classification problems, the classification error (CE) is defined as:

$$E_C = \|\mathbf{Y}_d - \text{round}(\mathbf{H}\mathbf{W})\|_2^2 \quad (5)$$

where $\text{round}(\bullet)$ is the round operation. Both the RE and CE are adopted into our algorithm, which is also called as the double error criterion. In order to decrease the computational complexity, a recursive algorithm is introduced to obtain the updating \mathbf{W} and the double errors. The further details can be referred to the literature [2].

For the recursive orthogonal least square (ROLS) algorithm in literature [2], the controlling parameter must be given beforehand, which can cause that the finally selected hidden centers are the optimal combinations only matching the pre-given controlling parameter. Generally, the controlling parameter is a function of many relative factors, and it is difficult to solve using the traditional methods. Therefore, to solve the optimal controlling parameter matching the currently selected hidden centers, the use of the PSO, a relatively new population-based evolutionary computation technique [3], is here preferred. To decrease the computational cost, assume that only one controlling parameter is used without any prior knowledge. And the corresponding fitness function f , in this paper, is defined as RE: $f = E_R$.

3 Experimental Results

3.1 Plant Species Identification Task

Plant species identification is a process in which each individual plant should be correctly assigned to a descending series of groups of related plants, as judged by their common characteristics. It is an important and essential task to correctly and quickly identify the plant species in collecting and preserving genetic resources, the discovery

of new species, plant resource surveys, plant species database management, etc. So far, although many plant taxonomy methods have been devised, such as molecular biological approaches, this time-consuming, troublesome task has mainly been carried out manually by botanists. Currently, automatic plant recognition from color images is one of most difficult tasks in computer vision because of a lack of proper models or representations for plant. Thus, the work of research and development in computer-aided plant species identification from color images is still in its infancy.

Bark is the outer protective coating of the trunk and branches of trees and shrubs, which includes all the tissues outside of the vascular cambium. The appearance of a bark depends on the type of cork cells produced by the cork cambium, the relative amount of cambial products, and the amount of secondary conducting tissue (phloem). The varied texture and thickness of bark are often functions of the environment in which the tree grows. The variation in the structure of bark often gives a tree its characteristic appearance, for example, the basswood's bark is brown/gray with deep vertical fissures and flat ridges, a crab apple's bark is reddish/brown, shallow fissures with broad flat topped scaly ridges, etc. A forester can recognize the species of trees by the differences in their bark either externally or by cutting a small slash to examine the inner structure. So, bark is a useful diagnostic feature for plant classification.

In our work, a bark image database is used in the following experiment, which was collected and built by ourselves in our lab. This database includes 22 species of different plants. Each species includes 20 leaves images, 10 of which are used as training samples. There are totally 440 images with the database. A subset of the images is shown in Fig.3.

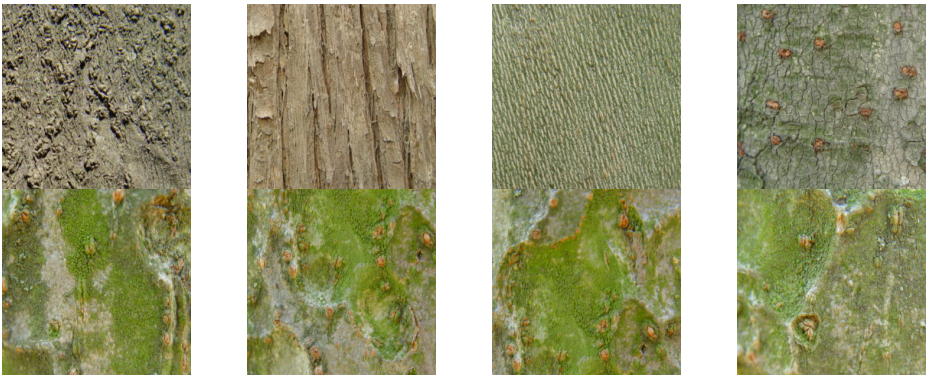


Fig. 3. A Subset of the Bark Texture Images

The Gabor filter has been widely adopted to extract texture features from images, and has been shown to be very efficient in doing so [4]. For each test image, five scales ($M = 5$) and six orientations ($N = 6$) were used. So there are 60 features for each test image, which is also the input vector of RBPNN.

The performance of RBPNN classifier is shown in Table 1. As observed from Table 1, the best recognition rate obtained by RBPNN Optimized by MMCH-HO method is 92.32%, respectively while the best recognition performance obtained by optimized RBFNN is 91.20%. And the optimized structure of RBPNN is simpler than the one of the optimized RBFNN, which indicate that the optimized RBPNN will have better generalization performance than the RBFNN.

Table 1. Performance Comparison of Different Classifiers for The Bark Texture Images

Classifiers	Classification Accuracy		Number of Hidden Neurons
	Training	Testing	
RBPNN (Optimized by MMCH-HO method)	98.81	92.32	83
RBFNN (Optimized by MMCH-HO method)	97.66	91.20	97
RBPNN (Not Optimized)	97.37	90.94	220
RBFNN (Not Optimized)	98.10	88.93	220

3.2 Palmprint Recognition Task

Then, to further demonstrate the efficiency of our proposed algorithms, we applied this algorithm to palmprint recognition. For a typical palmprint recognition system based on a neural network, firstly, some significant features are extracted in order to reduce data dimension and computational burden. Then, the recognition system is performed by neural networks. So, in this paper, we use a winner-take-all (WTA) network for independent component analysis (WTA-ICA) to extract features of palmprint images [5, 6]. According to literatures [5, 6], there are two types of implementation architectures for ICA in the image recognition task. The first architecture treats images as random variables and pixels as observations, i.e., each row of the input data matrix denotes an image, and its goal is to find a set of statistically independent basis images. While the other architecture utilizes pixels as random variables and images as observations, i.e., each column of the input data denotes an image, and its goal is to find a representation in which all coefficients are statistically independent. The detail of this algorithm can refer to literatures [5, 6].

We used the Hong Kong Polytechnic University (PolyU) palmprint database, available from <http://www.comp.polyu.edu.hk/~biometrics>, to verify our RBPNN algorithm. This database includes 600 palmprint images with the size of 128×128 from 100 individuals, with 6 images from each (as shown in Fig. 4). In all cases, three training images per person (thus 300 total training images) were randomly taken for training, and the remaining three images (300 total images) are taken for testing. To reduce the computational cost, each image is scaled to the size of 64×64 . Thus, the training set is a matrix with 300×4096 . During PCA preprocessing, compared with some different classifiers (BPNN, RBFNN, RBPNN), when the number of the first k PCs ($k > 85$), the recognition performance for all the classifiers have no remarkable

improvements. So the dimension of the training set is firstly reduced to 85 by PCA. Then using WTA-ICA architectures, 85 basis vectors (features) of palmprint images are extracted, i.e., all the ICs are used.

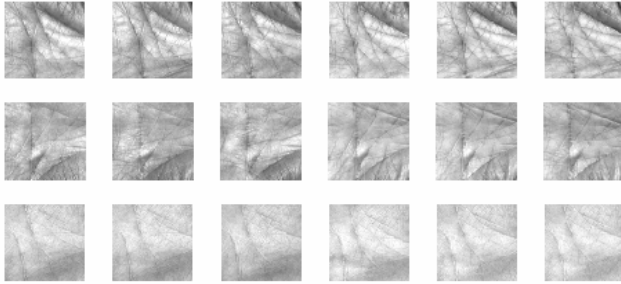


Fig. 4. Some palmprint image samples from PolyU palmprint database

Adopting the RBPNN to implement palmprint recognition task, all the 300 training samples are selected as the first hidden centers, the number of the second hidden neurons and the output layer neurons is set as 100, respectively. The controlling parameter needs to be set manually. When the controlling parameter is set to 0.211 and 0.196, respectively, which are determined by manually modifying and repeated experiments, the best recognition rates of the test samples corresponding to two WTA-ICA architectures are respectively 98.33% and 97.67%. With the same training and test samples, the RBFNN can achieve the highest recognition rates of 97.67% and 97.33% when the controlling parameter is set to 0.185 and 0.204, respectively.

In order to prune the RBPNN, the MVCH-HO method was also used to optimize the structure of the RBPNN. As a result, the number of the selected hidden centers of the first hidden layer is reduced from 300 to 116, and the recognition rates for the test samples corresponding to each WTA-ICA architecture are 98.33% and 99.33% (listed in Table 2), the corresponding controlling parameters are 0.232 and 0.238, respectively. Clearly, the optimized RBPNN structure can achieve a much better recognition performance. Compared with the RBPNN, with the same training and test data, by applying the MVCH-HO method to the RBFNN, the number of the selected hidden centers of the first hidden layer is reduced from 300 to 136, the maximum recognition rates of the RBFNN are respectively 97.33% and 98.33% (listed in Table 2), the corresponding controlling parameters are 0.203 and 0.212, respectively. Thus, it can be seen that the recognition rate of the RBPNN is higher than both that of the RBFNN. And WTA-ICA architecture II outperforms WTA-ICA architecture I in classification.

We also selected other three often used classifiers in palmprint recognition problem: BPNN, cosine measure classifier and Euclidean distance classifier to implement this task, the experimental results were also shown in Table 2. Obviously, these classifiers have lower recognition performances than the RBPNN, even the RBFNN.

Table 2. Classification Performance Comparison For The Palmprint Recognition

classifiers	WTA-ICA I	WTA-ICA II
RBPNN (Optimized by MMCH-HO method)	98.33	99.33
RBFNN (Optimized by MMCH-HO method)	97.33	98.33
RBPNN (Not Optimized)	97.67	98.33
RBFNN (Not Optimized)	97.33	97.67
BPNN	96.67	97.33
Cosine Measure	95.33	96.67
Euclidean Distance	93.33	94.33

4 Conclusions

This paper proposes a novel MMCH-HO approach to entirely optimize the structure of radial basis probabilistic neural networks (RBPNN), which is a new improvement strategy based on the original ROLS method [2]. The advantage of the proposed approach is that the structure of the RBPNN can be heuristically initialized; moreover, both the hidden centers and the controlling parameter can be entirely and simultaneously optimized. The experimental results show that our proposed MMCH-HO algorithm is feasible and efficient to optimize the RBPNN.

Acknowledgments. This work was supported by the grants of the National Science Foundation of China, No. 60805021, the China Postdoctoral Science Foundation (No.20060390180 and 200801231), and the grants of Natural Science Foundation of Fujian Province of China (No.A0740001 and A0810010).

References

1. Huang, D.S.: Radial Basis Probabilistic Neural Networks: Model and Application. *International Journal of Pattern Recognition and Artificial Intelligence* 13(7), 1083–1101 (1999)
2. Huang, D.S., Zhao, W.B.: Determining the Centers of Radial Basis Probabilistic Neural Networks by Recursive Orthogonal Least Square Algorithms. *Applied Mathematics and Computation* 162(1), 461–473 (2005)
3. Ioan, C.T.: The Particle Swarm Optimization Algorithm: Convergence Analysis and Parameter Selection. *Information Processing Letters* 85(6), 317–325 (2003)
4. Jafari-Khouzani, K., Soltanian-Zadeh, H.: Rotation-Invariant Multiresolution Texture Analysis using Radon and Wavelet Transforms. *IEEE Transactions on Image Processing* 14(6), 783–795 (2005)
5. Shang, L., Huang, D.S., Du, J.X., Zheng, C.H.: Palmprint Recognition Using FastICA Algorithm and Radial Basis Probabilistic Neural Network. *Neurocomputing* 69, 1782–1786 (2006)
6. Shang, L., Huang, D.S., Du, J.X., Huang, Z.K.: Palmprint Recognition Using ICA Based on Winner-Take-All Network and Radial Basis Probabilistic Neural Network. In: Wang, J., Yi, Z., Žurada, J.M., Lu, B.-L., Yin, H. (eds.) *ISNN 2006*. LNCS, vol. 3972, pp. 216–221. Springer, Heidelberg (2006)

Nonlinear Component Analysis for Large-Scale Data Set Using Fixed-Point Algorithm

Weiya Shi^{1,2} and Yue-Fei Guo²

¹ School of Information Science and Engineering, Henan University of Technology,
Zhengzhou 450002, China

² School of Computer Science, Fudan University,
Shanghai 200032, China
{wys, yfguo}@fudan.edu.cn

Abstract. Nonlinear component analysis is a popular nonlinear feature extraction method. It generally uses eigen-decomposition technique to extract the principal components. But the method is infeasible for large-scale data set because of the storage and computational problem. To overcome these disadvantages, an efficient iterative method of computing kernel principal components based on fixed-point algorithm is proposed. The kernel principle components can be iteratively computed without the eigen-decomposition. The space and time complexity of proposed method is reduced to $o(m)$ and $o(m^2)$, respectively, where m is the number of samples. More important, it still can be used even if traditional eigen-decomposition technique cannot be applied when faced with the extremely large-scale data set. The effectiveness of proposed method is validated from experimental results.

1 Introduction

Principal component analysis (PCA) is a classical method for feature extraction and dimension reduction [1]. It uses the dimensions with larger variances and neglects the less important components. Although PCA has been successfully used as a tool for dimension reduction, it does not work well in nonlinear data distribution. To generalize the principal component analysis in the situation of complex nonlinear data, kernel principal component analysis (KPCA) has been proposed which uses the kernel method [2]. Its main idea is to map the data set from the input space into high-dimensional feature space. Thus, the nonlinear components can be extracted using the traditional linear algorithm in feature space. The kernel trick is used to calculate the inner product between data set without knowing the explicit mapping function. The extracted nonlinear feature can be used in many complex applications, such as face recognition, image compression, et al.

The standard KPCA generally needs to eigen-decompose the Gram matrix [3], which is acquired using the kernel function. It must firstly store the Gram matrix of all data, which takes the space complexity of $O(m^2)$, where m is the number of data samples. In addition, it needs the time complexity of $O(m^3)$ to compute the kernel principal components. But traditional kernel function is based on the inner product of data vector,

the size of kernel matrix is too large when faced with the large-scale data set. This is infeasible for large-scale data set because of limited storage capacity.

In order to solve the problem of the large-scale data set, some methods have been proposed to compute kernel principal component. In general, these methods are classified into two category: the sampling and non-sampling based approaches. For the sampling based approaches, Zheng [4] proposed to partition the data set into several small-scale data set and handle them, respectively. Some representative data are chosen to approximate the original data set [5]. An EM algorithm is also proposed to extract the nonlinear components [6]. Williams et al. [7] and Smola [8] also proposed the similar approximation approach. But these methods will lose some information in the sampling process. Aside from that, it is time-consuming to search for the representative data. For the non-sampling based approaches, an iterative procedure is proposed to estimate the kernel principal components by kernelizing the generalize Hebbian algorithm [9]. But the convergence is slow and cannot be guaranteed. Gunter etc [10] enhanced the algorithm by incorporating a gain vector and an additional normalization term.

This paper proposes a method of computing Kernel principal components based on fixed-point algorithm (KPCA-F), which can effectively solve the problem of large scale data set. The method needs not to store the kernel matrix in advance and also avoids the eigen-decomposition. The iterative algorithm can find nonlinear eigenvector in just a few iterations. Its space and time complexity is reduced to $o(m)$ and $o(m^2)$, respectively. The effectiveness of proposed method is demonstrated by the experimental results on the artificial and real data set.

The rest of this paper is organized as follows: section 2 gives a short review of the Kernel PCA. Then, we describe the proposed method in section 3. The experimental evaluation of the proposed method is given in the section 4. Finally we conclude with a discussion.

2 Review of Kernel Principal Component Analysis

Let $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m)$ be the data matrix in input space, where $\mathbf{x}_i, i = 1, 2, \dots, m$, is a n -dimensional vector and m is the number of data samples. There exists a mapping function ϕ , which projects the data into high-dimensional (even infinite dimensional) Reproducing Kernel Hilbert Space (RKHS).

$$\begin{aligned} \phi : \mathbb{R}^n &\rightarrow F \\ x_i &\mapsto \phi(x_i) \end{aligned} \tag{1}$$

Using mapping function ϕ , we can get the data set $\Phi(\mathbf{X}) = (\phi(\mathbf{x}_1), \phi(\mathbf{x}_2), \dots, \phi(\mathbf{x}_m))$ in feature space. In practice, the mapping function ϕ needs not to be known explicitly and performed implicitly via kernel trick. A positive definite kernel function $\kappa(\cdot, \cdot)$ is used to calculate the dot product between mapped sample vectors, where $\kappa(\cdot, \cdot)$ is given by $\kappa(\cdot, \cdot) = \kappa(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$. The polynomial and Gaussian kernel are two widely used kernel function, given by:

$$\begin{cases} \kappa(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j)^d \\ \kappa(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) = \exp\left(-\frac{|\mathbf{x}_i - \mathbf{x}_j|^2}{2\sigma^2}\right) \end{cases} \tag{2}$$

Where d is the degree of the polynomial kernel function and σ is the width parameter of the Gaussian kernel function.

In mapping feature space, the covariance matrix is given as follows:

$$C = \frac{1}{m} \sum_{i=1}^m \Phi(\mathbf{x}_i) \Phi(\mathbf{x}_i)^T, \tag{3}$$

It accords with the following equation:

$$C\nu = \lambda\nu, \tag{4}$$

Where ν and λ are corresponding eigenvector and eigenvalue of covariance matrix. Because the solutions ν can be expanded using all the data set $\Phi(\mathbf{X}) = (\phi(\mathbf{x}_1), \phi(\mathbf{x}_2), \dots, \phi(\mathbf{x}_m))$ as:

$$\nu = \sum_{i=1}^m \alpha_i \phi(\mathbf{x}_i), \tag{5}$$

By substituting (3), (5) into (4), we can get the following formula:

$$K\alpha = m\lambda\alpha, \tag{6}$$

where α is span coefficient, K is Gram matrix denoted as $K = \Phi(\mathbf{X})^T \Phi(\mathbf{X}) = (\kappa_{ij})_{1 \leq i \leq m, 1 \leq j \leq m}$. The entry of Gram matrix is $\kappa_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j)$. It is proved [11] that the Gram matrix is positive semi-definite. To compute the kernel principal components, the traditional method is to diagonalize Gram matrix K using eigen-decomposition technique. After having achieved the eigenvector α , we can achieve the kernel principal components ν using (5). For a test sample \mathbf{x} , the nonlinear feature is:

$$(\nu, \phi(\mathbf{x})) = \sum_{i=1}^m \alpha_i (\phi(\mathbf{x}_i) \cdot \phi(\mathbf{x})) = \sum_{i=1}^m \alpha_i \kappa(\mathbf{x}_i, \mathbf{x}), \tag{7}$$

In the process of whole deduction, it is assumed that the data have zero mean, if not, we can get the centering matrix $\hat{K} = K - I_m K - K I_m + I_m K I_m$, where $I_m = (1/m)_{m \times m}$ [3].

3 Proposed Method

The fixed-point algorithm is famous numerical analysis method, which has been used in many application [12] [13]. In this paper, we extend the idea to deal with nonlinear data and find the leading eigenvectors effectively. In whole computation procedure, the kernel matrix needs not to store in advance and also avoids the eigen-decomposition. The nonlinear feature can be gotten in some iterations. More important, it still can be used even if traditional eigen-decomposition technique cannot be applied when faced with the extremely large-scale data set.

3.1 Nonlinear Component Analysis Using Fixed-Point Algorithm

Assuming $\tilde{\mathbf{x}}$ is the reconstructed vector of n -dimensional vector \mathbf{x} . The $d \times n$ matrix ω is the transformation to reduce dimension of vector \mathbf{x} from n -dimensional input space to d -dimensional feature space, and μ is the mean of all samples. Then, the mean squared error is given as:

$$\begin{aligned}
 MSE &= E[||\mathbf{x} - \tilde{\mathbf{x}}||^2] \\
 &= E[||\mathbf{x} - \omega\omega^T(\mathbf{x} - \mu) + \mu||^2] \\
 &= E[(I_{n \times n} - \omega\omega^T)(\mathbf{x} - \mu)||^2] \\
 &= E[(I_{n \times n} - \omega\omega^T)(\mathbf{x} - \mu)^T (I_{n \times n} - \omega\omega^T)(\mathbf{x} - \mu)] \\
 &= E[(\mathbf{x} - \mu)^T (\mathbf{I}_{n \times n} - \omega\omega^T)(\mathbf{x} - \mu)],
 \end{aligned}
 \tag{8}$$

We take the derivative of (8) with respect to ω ,

$$\frac{\partial}{\partial \omega} E = -2E[(\mathbf{x} - \mu)(\mathbf{x} - \mu)^T] \omega = \Sigma \omega
 \tag{9}$$

Where $\Sigma = E[(\mathbf{x} - \mu)(\mathbf{x} - \mu)^T]$ is the covariance of samples. In order to compute the values of ω , fixed-point algorithm can be used:

$$\begin{cases} \omega \leftarrow \Sigma \omega \\ \omega \leftarrow \text{orthonormalize}(\omega) \end{cases}
 \tag{10}$$

In KPCA, the main computation in (6) is to eigen-decompose the kernel matrix \mathbf{K} . When faced with large-scale data sets, the storage and computational problem makes it impossible. We could replace the covariance Σ of samples with the kernel matrix \mathbf{K} in (6). The formula is given by:

$$\begin{cases} \omega \leftarrow \mathbf{K} \omega \\ \omega \leftarrow \text{orthonormalize}(\omega) \end{cases}
 \tag{11}$$

Because the kernel matrix is positive semi-definite, it needs not to store in advance. We only need to store all m samples. In each iteration, the row of kernel matrix is computed between the input sample and stored samples. Then, the dot product between the acquired row and eigenvector is calculated. The Gram-Schmidt orthonormalization procedure can be used to orthonormalize the eigenvector.

The following formula is used as the convergence criteria:

$$\text{mean}(\text{abs}(\tilde{\omega}_p - \omega_p)) < \varepsilon
 \tag{12}$$

where $\tilde{\omega}_p$ is the new value of ω_p in each iteration, ε is the tolerance.

The computation process of the proposed method is shown in Algorithm 1:

Algorithm 1. Computation process of the proposed method**Require:** $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m)$ be the data matrix in input space

- 1: Choose the number p of principal component and set $p \leftarrow 1$
- 2: Initialize the eigenvector ω_p randomly
- 3: update ω_p using (11)
- 4: If ω_p has not convergence, go to step 3
- 5: Let $p \leftarrow p + 1$ and go to step 2
- 6: output the eigenvectors and eigenvalues
- 7: For a test sample x , the nonlinear feature is given by $\sum_{i=1}^M \alpha_i \kappa(\mathbf{x}_i, \mathbf{x})$

3.2 Computational Complexity of the Algorithm

In whole computation procedure, we needs not to store the kernel matrix and only store m samples. For each sample, we compute the similarity with all other samples using kernel function, the inner product is computed between the acquired vector and the estimated eigenvector. (11) is used to produce the final eigenvector. The space complexity is reduced from $o(m^2)$ to $o(m)$.

The computation complexity consists of updating process using Fixed-point algorithm and the Gram-Schmidt orthonormalization, which is approximative $o(m^2)$.

4 Experimental Results and Discussion

To demonstrate the effective of the proposed method, we use the standard KPCA and proposed method on two-dimensional toy problem. In addition, we use USPS data set to de-noise the noisy image using extracted kernel principal components. We also validate the feasibility of proposed method when traditional eigen-decomposition technique cannot be applied. In our experiments, the Gaussian kernel function $\kappa(\mathbf{x}, \mathbf{y}) = \exp(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2})$ is only used if it is explicitly stated(σ is kernel parameter which needs to be adjusted by cross-validation).

4.1 Toy Examples

We firstly use 2-dimensional toy problem to demonstrate the effectiveness of proposed method. The 200 2-dimensional data samples are generated, which x-values are uniformly distributed in $[-1, 1]$ and y-values are given by $y = x^2 + \eta$ (η is the normal noise with standard deviation 0.2). The polynomial kernel $\kappa(x, y) = (x^T y)^d$ (where $d = 2$ is the degree) is used.

The experiment results are given in Fig. 1. It gives contour lines of constant value of the first 3 principal components, where the gray values represent the feature value. From the result, the proposed method can get comparable performance with standard KPCA.

4.2 USPS Examples

We also test the proposed method on real-world data. The US postal Service (USPS) data set¹ is 256-dimensional handwritten digits '0' – '9'. It consists of 7291 training

¹ Available at <http://www.kernel-machines.org>

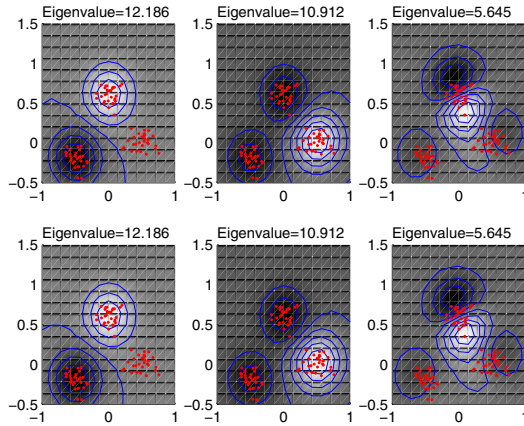


Fig. 1. Contour image of first 3 principal components obtained from the standard KPCA (the top row) and KPCA-F (the bottom row)

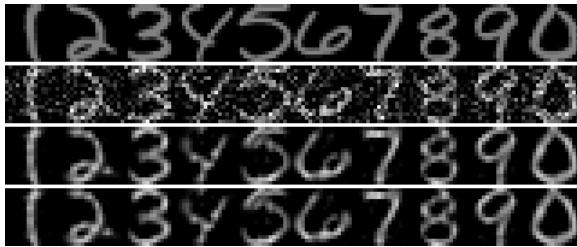


Fig. 2. De-noising results obtained by different methods. First row: original image; Second row: noisy image; Third row: de-noising result by batch KPCA; Fourth row: de-noising result by proposed method.

samples and 2007 testing samples. We extract some kernel principal components using KPCA-F and the standard KPCA, respectively. Then, the testing sample is projected on these extracted nonlinear principal components. Finally, the nearest neighbor classifier is used to classify the projecting testing sample.

We firstly use the extracted features to compute the pre-image and denoise the noisy image like [14]. The kernel parameter σ is set to $\sigma = dc$, where d is the image dimension and c equals to twice the average variance of data. The standard KPCA and the proposed method are trained with 3000 randomly choosing training samples. Then the testing samples were added additive Gaussian noise with zero mean and standard variance 0.5. After the non-linear feature is achieved, we reconstruct the noisy image using the first 64 kernel principal components. Fig. 2 gives the original testing image, noising image and the de-noising obtained by standard KPCA and the proposed method. It can be found that two methods achieve the similar de-noising image.

To further demonstrate the effective of proposed method, we use all the training samples to extract the nonlinear feature. The polynomial kernel $\kappa(x, y) = (x^T y)^d$ (where

Table 1. Error rate of 2007 testing sample using KPCA-F (having different degree d) when trained with all training samples

Number of components	Error rate for different degree				
	2	3	4	5	6
32	5.93	6.78	8.42	9.97	10.91
64	5.63	6.18	6.88	7.82	9.57
128	5.78	6.03	6.93	7.22	8.67
256	5.63	6.28	6.43	6.98	7.72
512	5.56	6.03	6.23	6.58	7.57

d is the degree) is used to compute the Gram matrix. Because the size of Gram matrix is 7291×7291 , it is impossible for standard KPCA algorithm to run. But the proposed method still works well. We extract the first leading principal components using the proposed method, and the testing sample is projected on these extracted principal components. The nearest neighbor classifier is used to classify the projecting testing sample. Table 1 gives the error rate of testing sample of USPS data set. We can see that the proposed method can achieve the classified performance even the eigen-decomposition technique cannot work out.

5 Conclusion

An efficient nonlinear component analysis for large-scale data set is proposed, which is based on fixed-point algorithm. Compared to other related methods, the proposed method is different in the following aspects: (a). needs not to store the kernel matrix in advance (b). avoids the eigen-decomposition. (c) The proposed method can be easily implemented; while many other methods are more complicated. The iterative algorithm can find nonlinear eigenvector in just a few iterations. Its space and time complexity is reduced to $o(m)$ and $o(m^2)$, respectively. The effectiveness of proposed methods is demonstrated by the experimental results on the artificial and real data set.

Acknowledgment

This work was supported in part by a grant from the National Natural Science Foundation of China (No. 60875003), Shanghai Municipal RND Foundation (No. 08511500902), National High Technology Research and Development Program of China (No. 2007AA01Z176).

References

1. Fukunaga, K.: Introduction to Statistical Pattern Recognition. Academic Press, London (1990)
2. Scholkopf, B., Smola, A.: Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond. MIT Press, Cambridge (2002)

3. Scholkopf, B., Smola, A., Muller, K.R.: Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Computation* 10, 1299–1319 (1998)
4. Zheng, W.M., Zou, C.R., Zhao, L.: An Improved Algorithm for Kernel Principal Components Analysis. *Neural Processing Letters* 22, 49–56 (2005)
5. France, V., Hlavac, V.: Greedy Algorithm for a Training Set Reduction in the Kernel Methods. In: *IEEE International Conference on Computer Analysis of Images and Patterns*, pp. 426–433 (2003)
6. Rosipal, R., Girolami, M.: An Expectative-maximization Approach to Nonlinear Component Analysis. *Neural Computation* 13, 505–510 (2001)
7. Williams, C., Seeger, M.: Using the Nystrom Method to Speed up Kernel Machine. In: *Advances in Neural Information Processing Systems* (2001)
8. Smola, A., Cristianini, N.: Sparse Greedy Matrix Approximation for Machine Learning. In: *International Conference on Machine Learning* (2000)
9. Kim, K.I., Franz, M.O., Scholkopf, B.: Iterative Kernel Principal Component Analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 27(9), 1351–1366 (2005)
10. Gunter, S., Schraudolph, N., Vishwanathan, S.V.N.: Fast iterative kernel Principal Component Analysis. *Journal of Machine Learning Research* 8, 1893–1918 (2007)
11. Shawe-Taylor, J., Cristianini, N.: *Kernel Methods for Pattern Analysis*, 3rd edn. Cambridge University Press, Cambridge (2004)
12. Hyvärinen, A., Oja, E.: A Fast Fixed-point algorithm for independent Component Analysis. *Neural computation* 9(7), 1483–1492 (1997)
13. Sharma, A., Paliwal, K.K.: Fast Principal Component Analysis using Fixed-point Algorithm. *Pattern Recognition Letters* 28, 1151–1155 (2007)
14. Mika, S., Scholkopf, B., Smola, A., Muller, K.R., Scholz, M., Ratsch, G.: Kernel PCA and de-noising in Feature Spaces. In: *Advances in Neural Information Processing Systems* (1998)

Optimal Reactive Power Dispatch Using Particle Swarms Optimization Algorithm Based Pareto Optimal Set

Yan Li¹, Pan-pan Jing¹, De-feng Hu¹, Bu-han Zhang¹, Cheng-xiong Mao¹,
Xin-bo Ruan¹, Xiao-yang Miao², and De-feng Chang²

¹ Electric Power Security and High Efficiency Lab,

Huazhong University of Science and Technology, Wuhan 430074, China

² Xinxiang Electric Power Supply Corporation of Henan Electric Power Company,
Xinxiang 453000, China

Abstract. An improved particle swarms optimization algorithm based on Pareto Optimal set is proposed to optimize the reactive power in power system, which is a multiple objectives optimization problem. The proposed algorithm develops the new fitness assignment and random inertia weight strategy, problem-specific linkages can be learned by examining a randomly chosen collection of points in the search space, the improved algorithm also has the ability to avoid getting trapped in local optima due to prematurity, applying it to the calculation of the power systems of IEEE6-bus and IEEE14-bus, the calculation results prove its effectiveness.

Keywords: Reactive power optimization, Pareto Optimal set, Fitness assignment, Random inertia weight strategy.

1 Introduction

Reactive power optimization is a multi-variable multi-restriction mixed non-linear optimization problem[1,2], furthermore, it involves controlling of continuous and discrete variables, which raises its complexity. The objective of reactive power optimization in power system is to control voltage quality and reduce the active power loss with the regulation of the voltage level of the generators, the taps of the transformer with OLTC and the switchable shunt capacitor/reactor groups. Therefore, it is significant to study reactive power optimization in order to ensure the security and reliability of power systems. With the development of the distributed generation technology, reactive power optimization problem is becoming more difficult, the study of effective optimization methods is an urgent task.

A lot of researches on reactive power optimization have been conducted, different algorithms have been developed, such as traditional methods including linear programming, non-linear programming, quadratic programming and dynamic programming, these methods can not ensure the acquirement of the whole optimization, with the development of the computer technology and artificial intelligence, Genetic Algorithm (GA) [3], Tabu search algorithm (TS) [4], and many evolutionary algorithms (EAs) have been proposed which have more or less success in solving various nonlinear engineering optimization problems. Among them, the best one is considered to be the

particle swarm optimization (PSO). It was first proposed by Kennedy and Eberhart[5-8], who derived the idea from the research on the bird preying. PSO has the characteristics of rapid convergence, high efficiency and strong robustness in multidimensional spaces and dynamic objects optimization, so it has been widely used in power economic dispatch, state estimation, reactive power optimization and voltage control[9-11]. However, it is difficult for PSO to solve reactive power optimization problems due to the multiple objectives and boundary restriction of the variables. In this paper, the random inertia weight strategy is applied to the initialization and flight of the particle, and specific linkages are incorporated to the searching process, which forms a new improved PSO to resolve actual reactive power optimization, the proposed algorithm is validated by the calculation and analysis of IEEE standard samples.

2 Particle Swarm Optimization

PSO is originated from the research of food hunting behaviors of birds, their behaviors are unpredictable but always consistent as a whole, with individuals keeping the most suitable distance. Every swarm of PSO is a solution in the solution space, it adjusts its flight according to its own and its companion's flying experience, the best position in the course of flight of each swarm is the best solution that is found by the swarm. Obviously, each swarm of PSO can be considered as a point in the solution space. If the scale of swarm is N , then the position of the particle is expressed as $X_i = (x_{i1}, x_{i2}, \dots, x_{in})$. The "best" position passed by the particle is expressed as $P_i = (p_{i1}, p_{i2}, \dots, p_{in})$. The speed is expressed with $V_i = (v_{i1}, v_{i2}, \dots, v_{in})$. The index of the position of the "best" particle of the swarm is expressed with $P_g = (p_{g1}, p_{g2}, \dots, p_{gn})$. Therefore, swarm i will update its own speed and position according to the following equations:

$$v_{id}^{k+1} = \omega v_{id}^k + c_1 r_1 (p_{id} - x_{id}) + c_2 r_2 (p_{gd} - x_{id}) \quad (1)$$

$$x_{id}^{k+1} = x_{id}^k + v_{id}^{k+1} \quad (2)$$

Where, $d = 1, 2, \dots, n$; k is the current generation; ω is the inertia weight factor; c_1 and c_2 are learning factors; r_1 and r_2 are two random numbers within the range $[0,1]$.

The equations consist of three parts, the first part is the former speed of the swarm, which shows the present state of the swarm; the second part is the cognition modal, which expresses the thought of the swarm itself, the third part is the social modal. The three parts together determine the space searching ability. The first part has the ability to balance the whole and search a local part, the second part causes the swarm to have a strong ability searching the whole and avoiding local minimum, the third part reflects the information sharing among the swarms. Under the influence of the three parts, the swarm can reach an effective and best position.

2.1 Pareto Optimal Set

The multi objective optimization problem can be expressed as follows:

$$\text{Min } y = f(x) = (f_1(x), f_2(x), \dots, f_n(x))$$

$$\text{s.t. } g_i(x) \leq 0.$$

Where $x \in R^m$ is the decision vector, $y \in R^n$ is the objective vector, $f_i(x)$ ($i=1,2,\dots,n$) is the objective function, and $g_i(x)$ is the system constraint. In most cases, the objective functions may conflict with each other, this may cause some multi objective optimization problems not to have the unique best global solution. However, there exists a solution that can not be further optimized for one or several objective functions and cannot be further worsened for other objective functions. This solution is called Pareto Optimal [7].

Definition 1. Let X^* be a point in the search space, it is a Pareto Optimal if there doesn't exist i (in the search space) which makes $f_i(x) < f_i(x^*)$ hold.

Definition 2. The set composed of all the Pareto Optimal is called Pareto Optimal Set, and is also called Acceptable Set or Effective Set.

The objective vectors corresponding Pareto Optimal are called non-dominator objective vectors. All the non-dominator objective vectors make up Pareto Front of a multi objective problem.

2.2 The Search of Pareto Optimal Set

This paper proposes the PSO assessed and chosen by the best solution and applied it to the search of Pareto Optimal Set in multi objective optimization problems. First, the algorithm initializes a particle swarm in the dominant vectors space. Then, the PSO directs the flight of swarm in the dominant vectors space together with each objective function in multi objective optimization problems, which causes the swarm to fall into the Pareto Optimal Set. Reflected in the space of objective function, the swarm will fall into Pareto Front.

The algorithm is performed as follows: First, find out the global best solution $g_{best}[i]$ ($i = 1,2, \dots N$, the number of the objective functions) and the best individual solution $p_{best}[i,j]$ ($j= 1,2, \dots N$, the number of the particles) in each swarm using each objective function in the multi objective optimization problems. The variables corresponding to each $g_{best}[i]$ in the dominant vector space make up an area called quasisolution area. When the speed of each swarm is updated, the "average" of each $g_{best}[i]$ is used as the best global solution g_{best} . Each particle's $p_{best}[i,j]$ is determined through judging the dispersed degree of vectors $p_{best}[i,j]$ and $g_{best}[i]$ to choose the "average" of the $p_{best}[i,j]$ or choose randomly in the $p_{best}[i,j]$. In addition, when the position of each particle is updated, it should be decided whether the position of each particle is within the quasisolution area. If it is then remain the original value, otherwise update the current value. The execution of the proposed algorithm is introduced using a two-objective optimization problem. Let's consider the minimization of $f_1(x)$ and $f_2(x)$.

(1) Initialize the particle swarm: Designate the population size N , generate speed V_i and position X_i of each particle randomly.

(2) Evaluate the fitness of each particle: Obtain $Fitness1[i]$ and $Fitness2[i]$ by using the two objective functions $f_1(x)$ and $f_2(x)$.

For $i = 1$ to N

$$Fitness1[i] = f_1(x[i]); \quad Fitness2[i] = f_2(x[i]);$$

Next i

(3) Calculate the best individual solutions $p_{best}[1, i]$ and $p_{best}[2, i]$.

For $i = 1$ to N

$$p_{best}[1, i] \leftarrow f_1(x) \quad p_{best}[2, i] \leftarrow f_2(x);$$

Next i

(4) Calculate the best global solutions $g_{best}[1]$ and $g_{best}[2]$.

$$g_{best}[1] \leftarrow f_1(x) \quad g_{best}[2] \leftarrow f_2(x) ;$$

(5) Calculate the ‘‘average’’ of the two best global solutions g_{best} and their distances dg_{best} :

$$g_{best} = Average(g_{best}[1], g_{best}[2]);$$

$$dg_{best} = Distance(g_{best}[1], g_{best}[2]);$$

(6) Calculate the distance $dp_{best}[i]$ between $p_{best}[1, i]$ and $p_{best}[2, i]$.

For $i = 1$ to N ;

$$dp_{best}[i] = Distance(p_{best}[1, i], p_{best}[2, i]);$$

Next i

(7) Calculate the best individual solution $p_{best}[i]$, which is used to update the speed v_i and position x_i of each particle:

For $i = 1$ to N

If ($dp_{best}[i] < dg_{best}$)

$$p_{best}[i] = Randselect(p_{best}[1, i], p_{best}[2, i]);$$

Else

$$p_{best}[i] = Average(p_{best}[1, i], p_{best}[2, i]);$$

Next i

(8) Update the speed v_i of each particle using g_{best} and $p_{best}[i]$.

(9) Judge whether the position x_i is in the quasisolution area. If it is then remain the original value, otherwise perform the update.

(10) If the termination condition is achieved then stop, otherwise go to step (2).

2.3 The Fitness Modification

The proposed PSO is assumed that various problem specific linkages are available in a Linkage Matrix L , where $L[i, j]$ indicates the strength of the linkage between the i th and j th components of the representations of candidate solutions[5], these values are presumed to lie in the interval $[0, 1]$. Procedurally, one component i is first chosen randomly to be the first element of the subset U , and every other component j is chosen to be included in U with a probability directly proportional to $L[i, j]$, its linkage with the first chosen component of U .

The attractor (i.e. the position towards which attraction is experienced) is chosen separately, and depends on the subset of components chosen for update (based on linkage strengths, as mentioned above). If X is the particle whose position is to be updated at time t using the subset of components U , then the attractor Y is chosen to maximize the function

$$g(U, Y, X, t) = \frac{f(p_{best}(Y, t)) - f(X)}{\|Proj(p_{best}(Y, t), U) - Proj(X, U)\|} \quad (4)$$

assuming the fitness function is to be maximized, where $p_{best}(Y, t)$ is the best position discovered by particle Y until time t, $\| \cdot \|$ is the Euclidean norm, and $Proj(Z, U)$ is the projection of Z onto the positions in subset U.

Then equation should be in the form of equation (5):

$$v_{id}^{k+1} = \omega v_{id}^k + c_1 r_1 (p_{id} - x_{id}) + c_2 r_2 (p_{gd} - x_{id}) + c_3 r_3 (p_{best}(Y, t)_d - x_{id}) \quad (5)$$

c_3 is the learning factor corresponding to P_{best} , when $d \in U$, $0 < c_3 < 1$; otherwise $c_3 = 0$. The increase of c_3 and the decrease of c_1 , c_2 have a strong influence on X. r_3 is a random number in the range $[0, 1]$.

In order to satisfy the requirement of the Pareto Optimal to the multi objectives optimizations, the fitness modification is developed:

(1) define the Pareto set. The Pareto set D_{Pareto} is used to record the pareto optimal in order to evaluate the fitness in the iteration process.

(2) evaluation of each pareto optimal in D_{Pareto} . $D_i = \frac{n_i}{N + 1}$ (6)

Where, n_i is the number of the dominant particles of i, N is the total number of the particle swarm.

(3) fitness calculation. $f_j = \frac{1}{1 + \sum_{i(i>j)} D_i}$ (7)

$i \in D_{Pareto}$, $j \in D$, $k \succ l$ means that k dominant l .

2.4 Random Inertia Weight Strategy

The role of the random inertia weight factor ω is considered critical for the convergence, it is employed to control the influence of the previous history of the velocities on the current one, and accordingly, the inertia weighting function regulates the tradeoff between the global and local exploration abilities of the swarms. The proposed PSO develops a new random inertia weight strategy to get a better solution. Define K as equation (8):

$$k = \frac{f(t) - f(t-10)}{f(t-10)} \quad (8)$$

the random inertia weight factor ω changes with the different value of K, shown in equation (9):

$$\begin{cases} \omega = a_1 + r/2.0 & k \geq 0.05 \\ \omega = a_2 + r/2.0 & k < 0.05 \end{cases} \quad (9)$$

where, $a_1 > a_2 > 0$, the random parameter r is used to maintain the diversity of the population and is uniformly distributed within the range $[0, 1]$.

3 Static Models on Reactive Power Optimization

The static model on reactive power optimization is introduced as(10), (11) and (12), whose target function is the least power loss with the voltage and the transmission lines loading restriction.

$$\min P_{Loss} = \sum_{i=1}^{N_i} V_i \sum_{j \in h} V_j (G_{ij} \cos \theta_{ij} + B_{ij} \sin \theta_{ij}) \quad (10)$$

$$\min \Delta V = \sum_{i=1}^{N_D} \left(\frac{V_i - V_i^{spec}}{\Delta V_i^{max}} \right)^2 \quad (11)$$

$$\max \Delta S = \sum_{i \in N_D} (S_{lim} - S_{ini}) \quad (12)$$

its equation restriction functions are introduced as follows:

$$P_i = V_i \sum_{j=1}^{N_i} V_j (G_{ij} \cos \theta_{ij} + B_{ij} \sin \theta_{ij}) \quad i \in N_i, i \neq N_E$$

$$Q_i = V_i \sum_{j=1}^{N_i} V_j (G_{ij} \sin \theta_{ij} - B_{ij} \cos \theta_{ij}) \quad i \in N_{PQ}$$

its inequality restriction functions are introduced as follows:

$$V_{Gmin} \leq V_G \leq V_{Gmax} \quad V_{Lmin} \leq V_L \leq V_{Lmax} \quad Q_{Gmin} \leq Q_G \leq Q_{Gmax} \quad S_L \leq S_{Lmax}$$

$$K_T \in [K_{Tmin}, K_{Tmin+1}, \dots, K_{Tmax-1}, K_{Tmax}] \quad Q_C \in [Q_{Cmin}, Q_{Cmin+1}, \dots, Q_{Cmax-1}, Q_{Cmax}]$$

in which, N_i is the total number of the buses, N_E is the slack buses, V_i^{spec} represents the specific voltage of node i, usually, $V_i^{spec} = 1$; $\Delta V_i^{max} = V_i^{max} - V_i^{min}$; the control variables are $X_c \in \mathbb{R}^n$ and $X_c = [V_G, K_T, Q_C]$ which represent the voltage level of the generators, the taps of the transformer with OLTC and the switchable shunt capacitor/reactor groups; the state variables are $X_s \in \mathbb{R}^n$ and $X_s = [V_L, Q_G, P_{ref}]$ which represent the voltage of buses, reactive power export of the generators and power export of reference bus; N_{pq} is the aggregate of PQ buses; S_L is apparent power on the branches. Assign the linkage matrix value as equations (12):

$$\begin{cases} L(i, j) = U(0.5, 1) & k_1 = k_2 \\ L(i, j) = 0.01 + \frac{1}{e^{2|k_2 - k_1|}} & k_1 \neq k_2 \end{cases} \quad (12)$$

4 Algorithm Implementation

The calculation process of improved PSO algorithm proposed by the paper is described as follows:

Step 1) Import the dimension and the limit of the control variables; set the limit of the status variables; Initialize the Linkage Matrix L, fix the scale N of particle swarm,

maximal iteration time, the upper and lower limit of random inertia weight ω , learning factors c_1 , c_2 and the random number r_1 , r_2 are assigned as statement in section 1.1.

Step 2) Suppose the current iteration $t=1$, calculate the objective functions shown in equations (10),(11) and (12). Calculate the best individual solutions p_{best} and the best global solutions g_{best} .

Step 3) calculate $p_{best}(Y, t)$ as equation (4).

Step 4) According to section 1.4, assign c_3 , r_3 . Calculate each particle's speed shown in equation (5), if it's smaller than its maximum velocity, update the current position as equation (2), or suppose $v_{id}^{k+1} = v_{id\max}$, then update the current position as equation (2).

Step 5) Check and reinforce the solution bounds, if it exceeds the limit, it's assigned as the corresponding limit value.

Step 6) If $t > t_{\max}$, output the optimal solution, or the stopping criteria are not satisfied, $t=t+1$, if $t > 10$, adjust the random inertia weight factor ω stated in section 1.5.

Step 7) Calculate the objective functions shown in equations (10),(11) and (12), calculate the best individual solutions p_{best} and the best global solutions g_{best} , jumps to step 3.

5 Case Study

The paper calculates the IEEE-6 and IEEE-14 sample network, the scope of particle swarm is 30, the maximal iteration times are 100, c_1 and c_2 are set as 0.2, the initial value of c_3 is 0.1, increases to 0.25, $\varepsilon = 5 \times 10^{-5}$.

Table 1. Optimal results and different methods comparison in IEEE-6

	ΔS	P_{Loss}	ΔV	$p / \%$
Initial state	0.1654	0.1088	0.9504	45.00
$PSO^{(1)}$	0.4417	0.0995	0.6896	77.72
$PSO^{(2)}$	0.3583	0.0872	0.9743	54.45
$PSO^{(3)}$	0.2096	0.1104	0.5175	72.40
proposed PSO	0.4409	0.0876	0.5182	93.28

$PSO^{(1)}$ 、 $PSO^{(2)}$ 、 $PSO^{(3)}$ is the conventional PSO, their objective functions are equations(10)(11)(12) respectively. $p = 1.1e^{(\Delta S - P_{Loss} - \Delta V)} \times 100\%$ reflects the integrative optimal ability.

5.1 The System of IEEE-6

The IEEE-6 buses sample includes two generators, two transformers with OLTC and two buses (node 4 and 6) with reactive power devices. The total loads of the system are $135\text{MW} + j36\text{MVar}$. The voltages of the generators change continuously between 0.9 to 1.1 p.u.; the step of the adjustable taps of the transformers is 0.025, the step of the capacitor is 0.01.

Table 2. Pareto-Optimized control variables and state variables

control variables	T_{65}	T_{43}	V_{G1}	V_{G2}	Q_{C4}	Q_{C6}
Lower limit	0.9	0.9	1.0	1.0	0.0	0.0
Upper limit	1.1	1.1	1.1	1.15	0.5	0.055
Initial value	1.000	1.000	1.050	1.100	0.000	0.000
PSO	0.9321	0.9475	1.100	1.1413	0.050	0.055
state variables	Q_{G1}	Q_{G2}	V_3	V_4	V_5	V_6
Lower limit	-0.2	-0.2	0.9	0.9	0.9	0.9
Upper limit	1.0	1.0	1.1	1.1	1.1	1.1
Initial value	0.443	0.277	0.923	0.9380	0.9070	0.9220
PSO	0.4135	0.1378	1.024	0.9965	1.0184	0.9771

Table 3. Pareto-Optimized state variables in IEEE-14

bus	P_s	P_L	Q_L	bus	P_s	P_L	Q_L
1	0.000	0.000	0.000	8	0.600	0.000	0.000
2	0.400	0.000	0.000	9	0.000	0.295	0.166
3	0.000	0.942	0.190	10	0.000	0.295	0.058
4	0.000	0.578	0.239	11	0.000	0.135	0.058
5	0.000	0.476	0.016	12	0.000	0.361	0.116
6	0.550	0.000	0.000	13	0.000	0.235	0.158
7	0.000	0.000	0.000	14	0.000	0.149	0.100

Table 4. Optimal results and different methods comparison in IEEE14-bus

	P_{loss}	ΔS	ΔV	$p\%$
Initial state	0.1841	0.2654	0.5718	60.95
PSO ⁽¹⁾	0.1370	0.3298	0.5725	75.25
PSO ⁽²⁾	0.1462	0.4385	0.5643	83.72
PSO ⁽³⁾	0.1589	0.3916	0.5310	81.63
Proposed PSO	0.1375	0.4381	0.5074	89.45

5.2 The System of IEEE-14

The IEEE-14 bus sample includes two generators, three transformers with OLTC and two buses (node 4 and 6) with reactive power devices. The voltages of the generators change continuously between 0.9 to 1.1; the step of the adjustable taps of the transformers is 0.025, the step of the capacitor is 0.01.

From table 1 and 4, the integrative optimal ability ($p/\%$) of the proposed PSO is the best among the different optimization methods. Comparing to the initial state, P_{Loss} and ΔV are decreased obviously, the improved PSO is effective to the multi objective optimizations in reactive power dispatch.

6 Conclusion

This paper demonstrates an improved PSO which exploits the multi objective optimization problems, the search of Pareto Optimal Set and the corresponding fitness modification are discussed in detail, a new random inertia weight strategy is developed, the improved PSO algorithm is applied on reactive power optimization in power system, case study in IEEE-6 and IEEE-14 shows its effectiveness, the paper also compares the performance of the proposed PSO against other convention PSO, the proposed one gives us the best performance in the sense that it converges to a more optimal solution. The proposed PSO has broad prospects for further study on the distribution generation system.

Acknowledgements. The authors would like to acknowledge that this research project is supported by the Key Project of National Natural Science Foundation of China (50837003) and the National Basic Research Program of China (2009CB219702).

References

1. Zhang, H., Hang, L., Meng, F.: Reactive Power Optimization Based on Genetic Algorithm. In: 1998 International Conference on Power System Technology, pp. 1448–1453. Sciences Press, Beijing (1998)
2. Abido, M.A., Bakhshwain, J.M.: Optimal VAR Dispatch Using a Multiobjective Evolutionary Algorithm. *Int. J. Elect. Power Energy Syst.* 27, 13–20 (2005)
3. Ma, J., Lai, L., Yang, Y.: Application of Genetic Algorithm in Reactive Power Optimization. In: Proceeding of the CSEE, pp. 347–353. Electrical Press, Beijing (1995)
4. Wang, H., Xiong, X., Wu, Y.: Power System Reactive Power Optimization Based on Modified Tabu Search Algorithm. *Power System Technology* 1, 15–18 (2002)
5. Deepak, D., Chilukuri, K.: Particle Swarm Optimization with Adaptive Linkage Learning. In: Proceedings of 2004 International Conference on Evolutionary Computation, pp. 530–535. IEEE Press, New York (2004)
6. Ho, S.L., Yang, S., Ni, G., Edward, W.C., Wong, H.C.: A Particle Swarm Optimization-Based Method For Multiobjective Design Optimizations. *IEEE Transactions on Magnetics* 41, 1756–1759 (2005)
7. Eckart, Z., Lothar, T.: Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Approach. *IEEE Transactions On Evolutionary Computation* 3, 257–271 (1999)
8. Carlos, A., Coello, C., Maximino, S.L.: MOPSO: A Proposal for Multiple Objective Particle Swarm Optimization. *IEEE Transaction On Evolutionary Computation* 2, 1051–1056 (2002)
9. Yu, J., Zhao, B.: Improved Particle Swam Optimization Algorithm for Optimal Power Flow Problems. In: Proceedings of the CSU-EPSA, pp. 83–88. IEEE Press, New York (2005)
10. Natsuki, H., Hitoshi, I.: Particle Swarm Optimization with Gaussian mutation. In: Proceedings of the 2003 IEEE, pp. 72–79. IEEE Press, New York (2003)
11. Liu, Z., Ge, S., Yu, Y.: Optimal Reactive Power Dispatch Using Chaotic Particle Swarm Optimization Algorithm. *Automation of Electric Power System* 29, 53–57 (2005)

A Robust Non-Line-Of-Sight Error Mitigation Method in Mobile Position Location

Sumei Chen, Ju Liu, and Lin Xue

School of Information Science and Engineering, Shandong University,
Jinan 250100, P.R. China

ju.liu@sdu.edu.cn

<http://202.194.26.100/liuju/index.htm>

Abstract. A novel non-line-of-sight (NLOS) error identification and range measurement reconstruction algorithm is proposed. First, a NLOS error identification technique is exploited to decide whether there is a NLOS path from the base station (BS) to the mobile station (MS); Second, a biased Kalman filter (KF) is used to mitigate NLOS error in the raw measurements; and then, two measuring points which have maximum residual (so-called the upper dead point and the lower dead point) are investigated to suppress the NLOS error as well as measurement noise. Under the assumption that the NLOS range measurements have been identified, we propose an orthogonal polynomial to smooth the range measurements estimated by KF and a range measurement reconstruction model to suppress NLOS error. Simulations verify its effectiveness to true range reconstruction as well as location tracking in NLOS environment.

Keywords: NLOS, residual, Kalman filter, orthogonal polynomial.

1 Introduction

A conventional method for locating a mobile station requires the measurements of the time of arrival (TOA) from at least three participating BSs. In the absence of any measurement error, the location of the MS can be unambiguously determined by the intersection of the circular curves. When the TOA measurements are corrupted by noise, the coordinates of the MS can be determined by finding the solution in a least-square sense [1]. However, due to reflection and diffraction, the direct path from the MS to the BSs may be blocked and the signal may actually travel excess lengths on the order of hundreds of meters, which is always the case in a dense urban environment. This phenomenon, which has been identified as one of the major factors that affect the accuracy of range measurements, is referred to as the NLOS problem and will ultimately translate into a biased estimate of the mobile station's position.

To mitigate the impact of NLOS error, some techniques are proposed [2,3] and Wylie's method [3] is a classical one. In this paper, a KF based range reconstruction technique is proposed. KF is suited for the tracking method for its famous

performance in object tracking [4,5]. We focus on three parts: NLOS error identification, NLOS error mitigation by biased KF and true range reconstruction technique.

The rest of this paper is organized as follows: The concept of NLOS identification technique is presented in section 2. NLOS error mitigation using biased KF is introduced in section 3. The true range approximation method is discussed in section 4. Finally, our simulations and conclusions are presented in section 5 and section 6, respectively.

2 NLOS Identification

Taking the thermal receiver noise, signal characteristics and the NLOS excess path length error into account, the range measurement between the BSm, $m=1,2,\dots,M$ and MS is given by

$$r_m(t_i) = D_m(t_i) + los_m(t_i) + nlos_m(t_i) \tag{1}$$

where $m=1,\dots,M$ is the BSs index and $i=0,\dots,k-1$ is the time instant index. $D_m(t_i)$ is the real distance, $los_m(t_i)$ is the measurement error and $nlos_m(t_i)$ is the NLOS error.

When there is a direct signal propagation path between the MS and BSm, the range measurement at time t_i is corrupted only by the standard system measurement noise $los_m(t_i)$ and the NLOS error $nlos_m(t_i) \equiv 0$. The NLOS error can be deduced from the probability density function of the propagation delay between direct path and other paths. An exponential model has been investigated in [6],

$$P(\tau) = \begin{cases} \frac{1}{\tau_{rms}} e^{-\frac{\tau}{\tau_{rms}}} & \tau > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

τ is the NLOS propagation delay, $\tau_{rms} = T_1 d^\epsilon y$ is the root mean square delay spread, which has a lognormal distribution and depends on the environment parameters [7].

Without loss of generality, one can model the NLOS error as a random variable with approximately finite support over the real axis, $0 \leq nlos_m(t_i) \leq \beta_m$, and the standard measurement noise, $-\alpha_m \leq los_m(t_i) \leq \alpha_m$. The composite error is linear combination of $los_m(t_i)$ with $nlos_m(t_i)$, which has approximately finite support over $-\alpha_m \leq Composite_error_m(t_i) \leq \beta_m + \alpha_m$.

Range measurements in NLOS environment generally have a larger variance than that in LOS environment, especially when the MS is moving. According to the Nokia range error histogram [8], an implied standard deviation of the NLOS propagation error $\sim 409m$ indicates the mean and standard deviation of the range errors increase significantly in a NLOS environment where NLOS error dominates the measurement noise [6]. To reconstruct the NLOS range measurements, it is necessary to know which range measurements (if any) contain NLOS errors. The range measurements at each BS are first smoothed by N-th

order polynomial and use least square technique to solve the coefficients [6]. The smoothed range measurements are presented as

$$s(t_i) = \sum_{n=0}^{N-1} \hat{a}(n)t_i^n \quad (3)$$

where $\hat{a}(n)$ is the polynomial coefficients for smoothed measurements. From the assumption that $\sigma_m^2 = E\{\text{los}^2(t_i)\}$ and by calculating the deviation

$$\hat{\sigma}_m = \sqrt{\frac{1}{K} \sum_{i=1}^K [s(t_i) - r(t_i)]^2} \quad (4)$$

A hypothesis test is then performed for NLOS identification. Also, a residual analysis rank test is used to exclude some uncertainty about the hypothesis test results.

In the following section, we assume that the NLOS range measurements have been successfully discriminated from those in LOS environment, and then we use the biased KF to mitigate NLOS error in the raw measurements. Finally, the range reconstruction models are applied to approximate the true range between MS and BSs.

3 KF Based NLOS Error Mitigation

3.1 Introduction to KF

Kalman estimator is linear, unbiased, and has minimum variance. The base of KF consists two equations:

$$s(k+1) = As(k) + w(k) \quad (5)$$

$$z(k) = Gs(k) + v(k) \quad (6)$$

Equation (5) and (6) are called state transition equation and measurement equation, respectively. $s(k)$ is state vector, $z(k)$ is measurement vector, A is state transition matrix, G is measurement matrix, $w(k)$ and $v(k)$ are additive noise components with the covariance $R(k)$ and $Q(k)$.

These two equations establish the relation between the state at the k^{th} time instant and the measurements before and at this time instant. The iterative algorithm is shown as below:

$$\tilde{s}_k = A\hat{s}_{k-1} \quad (7)$$

$$\tilde{P}_k = A\hat{P}_{k-1}A^T + Q \quad (8)$$

$$e_k = z_k - G\tilde{s}_k \quad (9)$$

$$K_k = \tilde{P}_k G_k^T (G_k \tilde{P}_k G_k^T + R_k)^{-1} \quad (10)$$

$$\hat{s}_k = \tilde{s}_k + K_k e_k \quad (11)$$

$$\hat{P}_k = (1 - K_k G_k) \tilde{P}_k \tag{12}$$

where \tilde{s}_k and \hat{s}_k denote the prediction and estimate of the state vector at the k^{th} time instant, \tilde{P}_k and \hat{P}_k are predicted and estimated covariance of error, e_k is innovation and K_k is Kalman gain.

3.2 NLOS Error Mitigation by Biased KF

In this paper, a biased KF is applied. After NLOS identification, we use biased KF to track the raw measurements and suppress the NLOS error. The state transition matrix A , measurement matrix G and state vector $s(k)$ are as below:

$$A = \begin{bmatrix} 1 & \Delta & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & \beta \end{bmatrix} \tag{13}$$

$$G = [1 \ 0 \ 1] \tag{14}$$

$$s(k) = [r(k) \ \dot{r}(k) \ b(k)]^T \tag{15}$$

where Δ is the sampling period, $b(k)$ is the NLOS error, $r(k)$ and $\dot{r}(k)$ are TOA and its first-order derivative, respectively. α and β are determined experimentally.

In the process of iterative algorithm, based on the fact that the moving trajectory is continuous and gradual change, the estimate at the k^{th} time can be denoted as below:

$$r(k) = r(k - 1) + \Delta \cdot \dot{r}(k - 1) \tag{16}$$

Through iterative algorithm, the estimates of the state variables at each temporal point can be obtained according to the measurements at the sequential temporal points.

4 Range Reconstruction

4.1 The Range Reconstruction Model

Given a sufficient observation interval $\{t_i, i = 1, \dots, N\}$ at each BS, the range reconstructed in NLOS condition can be modeled as

$$R(t_i) = w_1[s(t_i) + \delta_1 - \alpha_m - \beta_m] + w_2[s(t_i) - \delta_2 + \alpha_m] \tag{17}$$

where $R(t_i)$ is the reconstructed range curve, δ_1 and δ_2 are positive values representing the maximum deviation from above and below the fitting curve, respectively. w_1 and w_2 are weight factors of the two maximum deviations, $s(t_i)$ is the least square fitting curve of range measurements.

Similarly, the range reconstructed in LOS condition can be modeled as

$$R(t_i) = w_1[s(t_i) + \delta_1 - \alpha_m] + w_2[s(t_i) - \delta_2 + \alpha_m] \tag{18}$$

4.2 NLOS Error Correction

After NLOS error mitigation by KF, we employ a range reconstruction technique based on the upper and lower dead points.

Given a sufficient observation time interval $[t_0, t_{K-1}]$ and independent range measurements, the composite error can reach its extreme $-\alpha_m$ and $\beta_m + \alpha_m$ at some instant t_1 and t_2 , respectively. Under this assumption, the true range measurements are reconstructed in three steps as follows:

First, we propose a fitting polynomial to smooth the estimates by KF at observation time interval $[t_0, t_{K-1}]$. A polynomial in a least-square sense

$$\begin{aligned} s(t_i) &= \text{span}\{\varphi_0(t_i), \varphi_1(t_i), \dots, \varphi_n(t_i)\} \\ &= a_0\varphi_0(t_i) + a_1\varphi_1(t_i) + \dots + a_n\varphi_n(t_i) \end{aligned} \tag{19}$$

is used to fit the date at each BSm and the coefficients $\{a_0, a_1, \dots, a_n\}$ are decided by

$$\{a_0, a_1, \dots, a_n\} = \underset{\{a_0, a_1, \dots, a_n\}}{\text{argmin}} \sum_{i=1}^K [s(t_i) - r(t_i)]^2 \tag{20}$$

In [6], the range measurements at each BS are smoothed by a N-th order polynomial (19) and the basis of fitting function is chosen to be $\{1, t, t^2, \dots, t^{N-1}\}$. But simulations show that the coefficient matrix of the normal equation is ill-conditioned. Here we used a recursive orthogonal polynomial to approximate the true range by

$$\begin{cases} \varphi_0(x) = 1 \\ \varphi_1(x) = (x - \alpha_0)\varphi_0(x) \\ \dots \\ \varphi_{k+1}(x) = (x - a_k)\varphi_k(x) - \beta_k\varphi_{k-1}(x) \end{cases}, k = 0, 1, \dots, m \tag{21}$$

and

$$\begin{cases} \alpha_0 = \frac{(x\varphi_0, \varphi_0)}{(\varphi_0, \varphi_0)} \\ \alpha_k = \frac{(x\varphi_k, \varphi_k)}{(\varphi_k, \varphi_k)} \\ \beta_k = \frac{(\varphi_k, \varphi_k)}{(\varphi_{k-1}, \varphi_{k-1})} \\ a_k = \frac{(f, \varphi_k)}{(\varphi_k, \varphi_k)} \end{cases} \tag{22}$$

where a_k is the coefficients of the orthogonal polynomial, $\varphi_{k+1}(x)$, $k = 0, 1, \dots, m - 1$ is the basis function and m is the fit-order of polynomial.

Second, we calculate the residual deviation from above and below the smoothed range measurements by $\delta_1 = \max_{t_i} r(t_i) - s(t_i)$ and $\delta_2 = \max_{t_i} s(t_i) - r(t_i)$.

Finally, we reconstruct the true range measurements by (17) when the path between the BS and MS is NLOS and (18) when it is LOS.

5 Simulation and Results

In this section, we present simulation examples to evaluate the performance of our models and algorithm described in the preceding sections. Four BSs are participant in the location procedure. BS1 is the serving BS and the MS is assumed to start from an initial position (-300,-100). The speed along the X-axis and Y-axis is a 10 m/s mean random variable with the same standard deviation of 2m, respectively. The measurement error is a zero mean Gaussian random variable with the standard deviation of 30m. The sampling period is 0.5s and 100 samples are taken. The NLOS error is given by (2) with environment parameters $T_1 = 1\mu s$, $\varepsilon = 0.5$ and the standard deviation $\sigma_\rho = 4dB$ in bad urban.

Fig.1 plots the range trajectories of different methods. Compared with the estimated results of biased KF without range reconstruction and Wylie’s method [3], our proposed method can approximate the real range curve better.

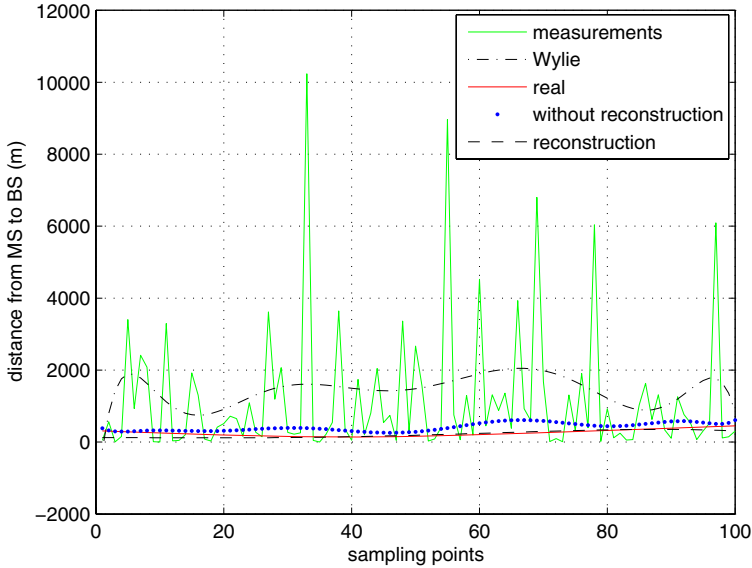


Fig. 1. Performance comparison between different methods

Fig.2 compares the mean-square estimation error (MEE) by using our proposed method to that of without using the reconstruction method. The mean-squared estimation error (MEE) is defined as the mean distance of the estimated mobile location to the true location. From this figure, we can see after range reconstruction, the location error decreases evidently and the performance improvement is obvious. Fig.3 shows the tracking trajectory of the mobile station. After range reconstruction, the track is close to the true MS’s position. On the contrary, track without reconstruction produces a comparative large deviation. Obviously, our method is better.

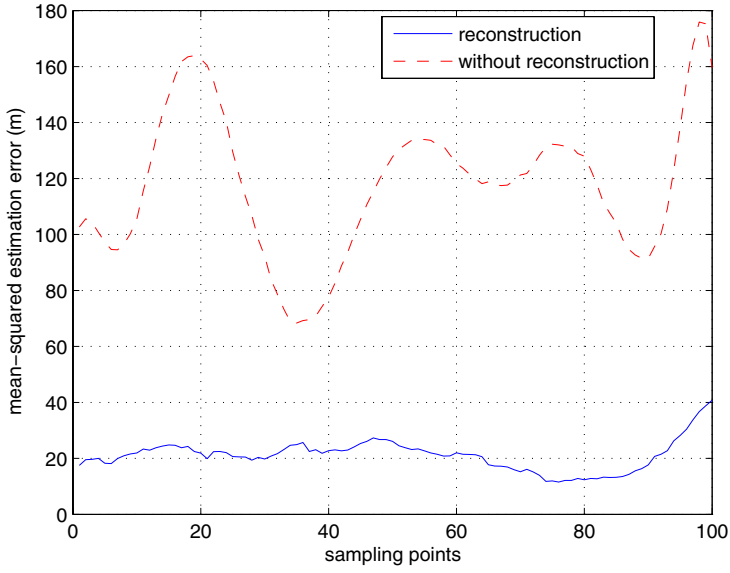


Fig. 2. MEE comparison

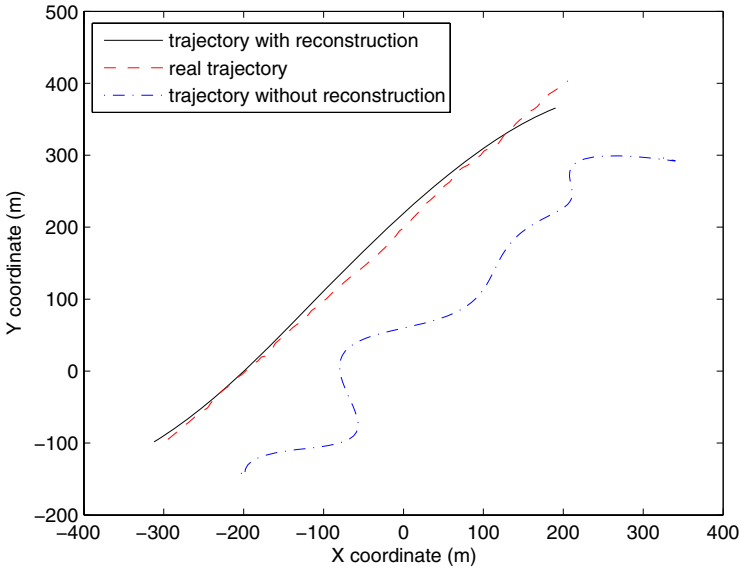


Fig. 3. Tracking trajectory

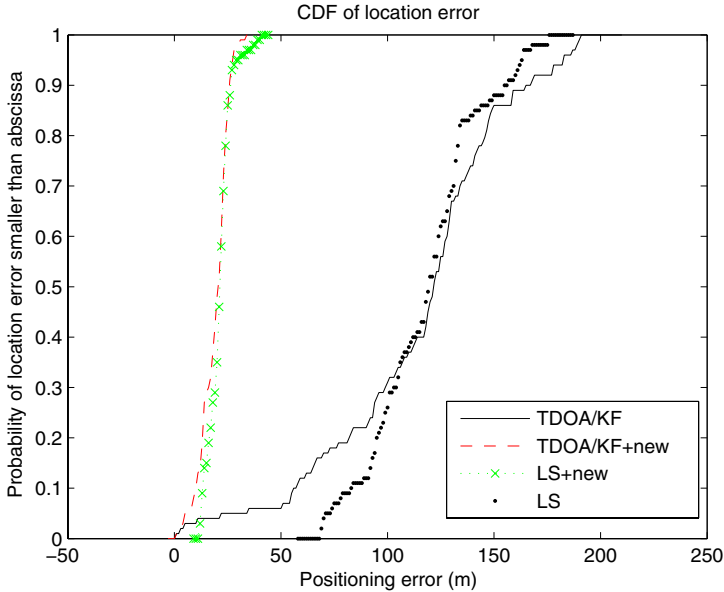


Fig. 4. Location error CDFs comparison

Fig.4 plots the CDFs (cumulative distribution function) of the location error of the proposed method and that of without range reconstruction using both TDOA linear KF estimator and LS estimator [1]. X-axis stands for location error in meter, Y-axis stands for the probability of location error smaller than the corresponding number in X-axis. In the legend, TDOA/KF and TDOA/KF+new denote TDOA location using linear KF without and with range reconstruction, respectively. Similarly, LS and LS+new denote TOA location using least square method without and with range reconstruction, respectively. It is evident that the performance of the proposed method is better than that of without range reconstruction, and our method can significantly improve the accuracy of mobile location. Also, we can see the performance of LS location and TDOA linear KF location is similar in the same condition. Thus our method suits both TDOA location and TOA location and the distinction between them is negligible. The pivotal matter depends on whether implementing range reconstruction.

6 Conclusions

In this paper, we present a new range reconstruction algorithm using orthogonal polynomial based on biased KF. After a NLOS propagation identification procedure in [2], a biased KF is used to smooth measurements and suppress NLOS error, and then, we can effectively approximate the true range based on residual analysis of the range measurements with respect to the orthogonal polynomial fitting curve. Simulation results indicate the new method can significantly decrease the mean-squared estimation error (MEE).

Acknowledgments. This work is supported by National Natural Science Foundation of China (605720105, 60872024), the Cultivation Fund of the Key Scientific and Technical Innovation Project (708059), the Program for New Century Excellent Talents (NCET-05-0582) in University, Natural Science Foundation of Shandong Province (No.Y2007G04), open research fund of National Mobile Communications Research Laboratory (W200802), and the State Key Lab. of Integrated Services Networks (ISN9-03).

References

1. Cheung, K.W., So, H.C., Ma, W.K., Chan, Y.T.: Least Squares Algorithms for Time-of-Arrival-Based Mobile Location. *IEEE Trans. Signal Process.* 52(4), 1121–1130 (2004)
2. Chen, P.C.: A Non-Line-of-Sight Error Mitigation Algorithm in Location Estimation. In: *Proc. IEEE Wireless Communications and Networking Conf. (WCNC)*, pp. 316–320 (1999)
3. Wylie, M.P., Holtzman, J.: The Non-Line of Sight Problem in Mobile Location Estimation. *Mobile Europe*, 827–831 (1995)
4. Najar, M., Vidal, J.: Kalman Tracking Based on Tdoa for Umts Mobile Location. In: *12th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 1 (2001)
5. Najar, M., Vidal, J.: Kalman Tracking for Mobile Location in Nlos Situations. In: *14th IEEE Proceedings on Personal, Indoor and Mobile Radio Communications*, vol. 3, pp. 2203–2207 (2003)
6. Lee, C.Y.: *Mobile Communications Engineering*, ch. 9. McGraw-Hill, New York (1993)
7. Greenstein, L.J., Erceg, V., Yeh, Y.S., Clark, M.V.: A New Path-Gain/Delay-Spread Propagation Model for Digital Cellular Channels. *IEEE Trans. On Vehicular Technology* 46, 477–484 (1997)
8. Silventoinen, M.I., Rantalainen, T.: Mobile Station Emergency Locating in GSM. In: *IEEE International Conference on Personal Wireless Communications*, India (1996)

Research on SSVEP-Based Controlling System of Multi-DoF Manipulator

Hui Shen, Li Zhao, Yan Bian, and Longteng Xiao

Department of Automation Engineering, Tianjin University of Technology and Education
Tianjin, P.C. 300222, China
shenhui19840730@163.com

Abstract. Steady State Visual Evoked Potential (SSVEP) rapidly becomes a practical signal of brain-computer interface system due to the advantage of high transmission rate and short training time. A SSVEP-Based controlling system of multi-dof manipulator is presented in this paper on the basis of virtual instruments. In this system, the ssvep-based electroencephalogram (EEG) was derived from scalp and then translated to several controlling commands of manipulator. In order to improve the performance of the system, the wavelet transform and Short-time Fourier Transform were used in signal processing. The experiment results have proved the effectiveness of the proposed method. The realization of the system can provide a new way to the using of robot-assisted in space based on BCI.

Keywords: Steady state visual evoked potential (SSVEP), Wavelet transform, LabVIEW, Short-time fourier transform.

1 Introduction

According to the stimulator, Visual Evoked Potential (VEP) can be divided into Transient Visual Evoked Potential (TVEP) and Steady State VEP Potential (SSVEP). The stimulating frequency of TVEP is often less than 4Hz, and the corresponding response disappears before the next stimulation. If the stimulating frequency is more than 6Hz, the responses are overlapped and emergence a Steady State VEP. The SSVEP is a non-invasive and practical input signal of brain-computer interface (BCI) because of its high information transfer rate and short training time.[1][2]

Based on the traditional BCI, a SSVEP-Based controlling system of multi-dof manipulator is presented in this paper on the basis of virtual instruments. In order to improve the performance of the system, the wavelet transform and Short-time Fourier Transform were used in signal processing, because the EEG is so weak and easy to be submerged in a variety of noises. The experiment results have proved the effectiveness of the proposed method. The communication of this system adopted wireless network and Multiple Input Multiple Output (MIMO) technology.

2 The SSVEP-Based Controlling System

A SSVEP-Based Controlling System of Multi-DoF Manipulator is presented in this paper. The system was consisted of EEG-evoked module, signal acquisition module, signal processing module and controlling exporting and external manipulator. Fig.1 shows the set-up of the system.

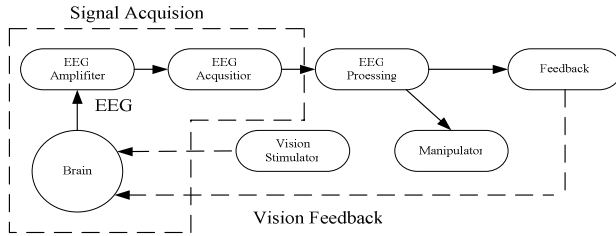


Fig. 1. The set-up of the system

2.1 The Hardware

The hardware system was composed of visual stimulation unit, digital EEG instrument, multi-function acquisition card and manipulator.

2.1.1 Visual Stimulation Unit and Acquisition System

The visual stimulation unit was composed of six LEDs flickering at corresponding frequency, the range of the flickering frequency was 8Hz~20Hz. The 9216sm digital EEG instrument was used to obtain in integral whole acquisition and A/D conversion, because it have better anti-interference capability. The data acquisition card worked in the synchronous model. The system used the NI PXI-6070E with high-performance data acquisition capacity since its highest sampling rate can reach up to 1.25M S/s, 12 bit resolution.

2.1.2 Manipulator

Fig.2 shows the manipulator of this system. It included a rotary joint and two translation joints. The rotary joint used direct current motor as drives, and translation joints used a stepping motor as drives, the other two rotary joints equipped with a resolution of 500p/n incremental rotary encoders, providing the half closed-loop feedback signal. The two translation joints were installed the trip-switch at the both ends. The scope was 0~256mm and there was a fixture at the end of the Manipulator. The Manipulator was controlled by a singlechip ATmega16.

The communication of this system adopted wireless network and Multiple Input Multiple Output (MIMO) technology which was put forward by Bell Labs in last century[3].It is a multi-antenna communications systems.

Since the feature of original EEG could not control the manipulator directly,it need a series switches re-coding with the feature in order to gain the control commands.

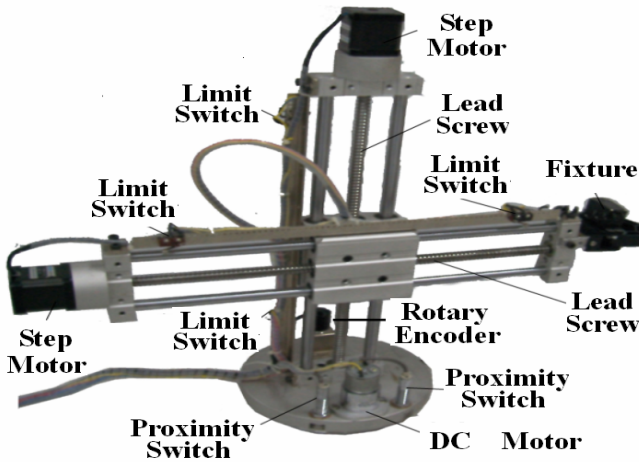


Fig. 2. Manipulator system

In the communication, the host machine and the manipulator executed the "one question and one answer" strictly. It would feedback data or null command depended on feedback. Serial baud rate was 19200bps,8 data bits, 1 stop bit and no parity bit.

2.2 The Software

The system was built on the virtual instrument platform, and designed by LabVIEW [4] program language. The Fig.3 shows human-machine interface of manipulator controlling system based on SSVEP.

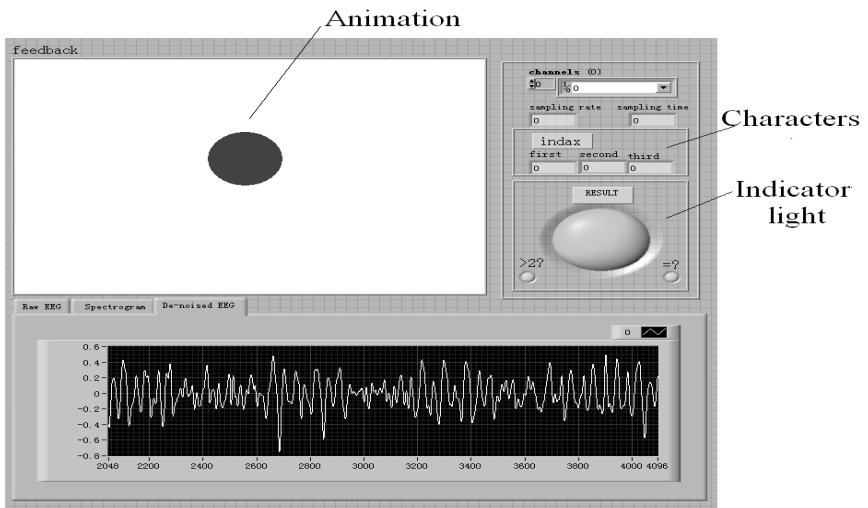


Fig. 3. Human-machine interface

Human-machine interface was composed of three functional areas: parameter setting module, graphic display module and feedback module. Signal acquisition channels, sampling frequency and sampling time could be set in parameter setting area. The graphic display module displayed the real-time EEG, de-noised EEG and EEG frequency spectrum for directly observing the EEG variation. In the feedback area, results were feed back in different forms such as indicator light, animation and characters. Characters can feed back the name of continuous three features extracted from EEG. In the indicator light area, the light turned red when the experiment failed. The left light indicated the amplitude of SSVEP whether more than the double mean value of EEG or not. The right light indicated the continuous three SSVEP extracting results. When the experiment finished effectively, the ball in the animation area can move the same direction of up, down, right and left as the manipulator do. The screen real-time feedback would make subjects more confident to control the manipulator better.

3 Data Processing

The amplitude of event-related potentials is about 2~10μV, which is smaller than the spontaneous EEG. The noises of SSVEP contain the spontaneous potential, the background noise of equipment, the interference is 50Hz city power and so on.

3.1 Wavelet De-noise

Function $\psi(t) \in L^2(\mathbb{R})$, if its Fourier transform $\psi(\omega)$ satisfies conditions:

$$C_\psi = \int_{\mathbb{R}} \frac{|\psi(\omega)|^2}{|\omega|} d\omega < \infty \tag{1}$$

$\psi(t)$ is called basic wavelet or mother wavelet function. And that

$$WT_f(a, \tau) = \frac{1}{a} \int_{\mathbb{R}} f(t) \psi^* \left(\frac{t - \tau}{a} \right) dt \tag{2}$$

is the continuous wavelet coefficients of mother wavelet and called wavelet transform of $\psi(t)$. Inverse wavelat transform formula is

$$f(t) = \frac{1}{C_\psi} \int_0^{+\infty} \frac{da}{a^2} \int_{-\infty}^{+\infty} WT_f(a, \tau) \psi_{a,\tau}(t) d\tau \tag{3}$$

Wavelet transform is the one that integrate signal with wavelet function which has well retractility in frequency domain and time domain. The signal is divided into different frequency band and time interval according to multi-resolution proposed by Mallat in the reference [5] in 1988. It can decompose the signal in different scales, and divide the signal to different sub-band so as to process the signal in different frequency bands.

This paper used wavelet soft-threshold de-noising[6] method. The de-noising was composed of three steps: (1) decomposing the original EEG. The wavelet selected sym2 and scale chose 3. (2) setting the threshold of every level coefficients. (3) doing signal reconstruction.

3.2 Short-Time Fourier Transform

SSVEP distributed in specific frequency, so it simplifies the EEG feature extraction algorithms. This system used short-time Fourier transform Every 4s, extracting a the signal feature by the principle if the extraction results were the same one for three times, then the feature was the issue of controlling orders.

4 Experiment Design and Analysis

SSVEP is raised by external stimuli, so subjects with almost no training will be able to achieve good effect. In our experiment, five subjects participated in the experiment.

In the experiment, subjects were seated in a comfortable chair, 50cm in front of the stimuli. They were asked horizontally watching the flickering LED. Electrodes were placed at O1 and O2 in line with international 10-20 system, and the ears were reference potential. Signal sampling rate was 512Hz. The average detection accuracy over all subjects is about 72%. The average information transfer rate is satisfied. Tab 1 shows the accuracy result of this system.

Table 1. Accuracy of the system

Result Subjects	Number of tests	Number of responses	Acuracy
HDF(BOY)	20	16	80%
YPX(GIRL)	20	14	70%
S H (GIRL)	20	15	75%
XLT(BOY)	20	13	65%
SPY(BOY)	20	14	70%
Average accuracy	72%		

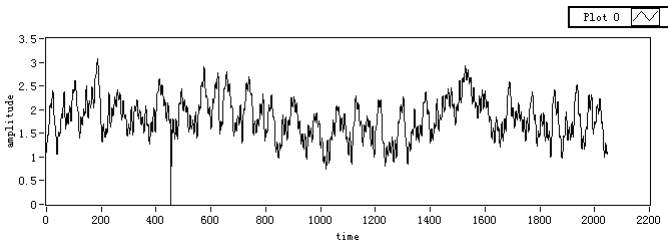


Fig. 4. The original EEG

Taking the stimuli frequency of 12Hz for example, 12Hz SSVEP represented the control command of down. Fig.4 shows the signal of original EEG. Wavelet denoise signal is shown in Fig.5. Fig.6 shows the frequency spectrum in the range of 6~40Hz.

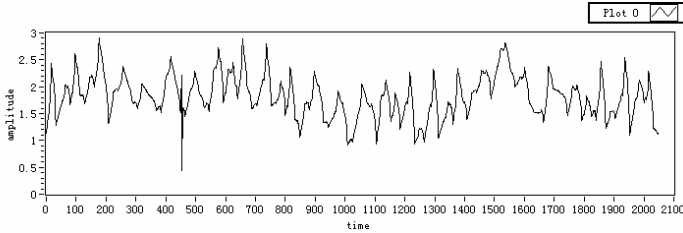


Fig. 5. The de-noised EEG

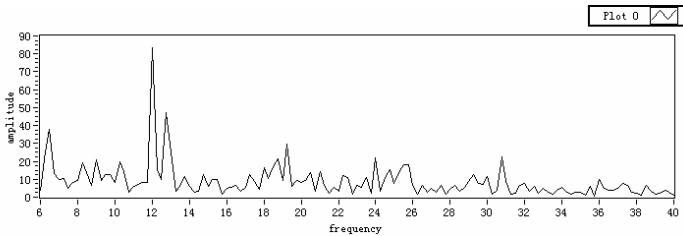


Fig. 6. Frequency spectrum

As can be seen from the frequency spectrum, this test was successful. Then the manipulator moved down, and in the animation feedback area, the ball moved down accordingly.

This system designs a real-time multi-DOF manipulator controlling system based on SSVEP, although real-time control have been achieved, there are also some defects need to improve. Generally, there are two important criterions to evaluate a BCI system, the first is accuracy, and the second is the speed. The BCI system can be improved by the following two methods: Improving signal-to-noise ratio by modifying the mother wavelet and implementing the effective method of pattern recognition to improve the speed of real-time communication system.

5 Conclusion

This paper designed a real time controlling system of multi-degree of freedom manipulator based on SSVEP. Through experimental results, the accuracy of the system was satisfied. It reflected the high-performance, short developing time and integration of hardware and software of visual instrument. The realization of the system provided a new way to using of robot-assisted in space based on BCI.

Acknowledgments. The work was supported by the Chinese National Program for High Technology Research and Development under the grant No.2007AA04Z254.

References

1. Wolpaw, J.R., Birbaumer, N., Pfurtscheller, D.J.G., Vaughan, T.M.: Brain-computer Interfaces for Communication and control. *Clin. Neurophysiol.* 113, 767–791 (2002)
2. Vaughan, T.M., Heetderks, W.J., Trejo, L.J., Rymer, W.Z., Weinrich, M., Moore, M.M., Kubler, A., Dobkin, B.H., Birbaumer, N., Donchin, E., Wolpaw, E.W., Wolpaw, J.R.C.: Brain-computerinterface Technology: A Review of the Second International Meeting. *IEEE Trans. Neural Syst. Rehabil. Eng.* 11, 94–109 (2003)
3. Zhao, L., Li, C., Cui, S.G.: Service Robot System Based on Brain-Computer Interface Technology. In: *The 3rd International Conference on Natural Computation*, Haikou, China (2007)
4. National Instruments Corporation. LabVIEW for ECG Signal Processing, <http://www.ni.com>
5. Mallat, S.C.: A Theory for Multi-resolution Signal Decomposition: The Wavelet Representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 11, 674–693 (1989)
6. Hazarika, N., Chen, J.Z., Tsoi, A.C.: Classification of EEG Signals Using the Wavelet Transform. *DSP*, 89–92 (1997)

Tracking Control of Robot Manipulators via Orthogonal Polynomials Neural Network

Hongwei Wang¹ and Shuanghe Yu²

¹ School of Electronic and Information Engineering, Dalian University of Technology

² School of Information Science and Technology, Dalian Maritime University
Dalian 116024, Liaoning, China

wanghw@dlut.edu.cn

Abstract. In this paper, an orthogonal functions neural network is used to achieve the control of nonlinear systems. The adaptive controller is constructed by using Chebyshev orthogonal polynomials neural network, which has advantages such as simple structure and fast convergence speed. The adaptive learning law of orthogonal neural network is derived to guarantee that the adaptive weight errors and tracking errors are bound by using Lyapunov stability theory. Simulation results are given for a two-link robot in the end of the paper, and the control scheme is validated.

Keywords: Chebyshev polynomials, Orthogonal neural network, Robot manipulators, Lyapunov stability theory.

1 Introduction

Generally, robot manipulators used as industrial automatic elements are known as the system with high nonlinearities, uncertainties and time-varying. If a controller of robot manipulator is designed, some factors should be considered, including the exact trajectory performance of reference input and the robustness for the existence of external disturbances. The conventional feedback controllers such as PID controller are commonly used in the industry field because their control architectures are very simple and easy to implement. However, when these conventional feedback controllers are directly applied to nonlinear systems, they suffer from the poor performance and low robustness due to the uncertainties and the external disturbances[1-3].

During the past decade, much researching effort has been put into the design of intelligent controllers using neural network. Recently, hybrid control methods containing neural network, fuzzy logic and other optimizing schemes have been attracted more and more attention[4-7]. Neural networks have used to adjust and optimize parameters of fuzzy controllers. However, backpropagation algorithm of neural network has the problems of local minimum, slow convergence speed, and difficulty in determination of the number of processing elements. In recent years, researchers have been tried to solve these problems or even to develop new structure of neural network[8,9]. They introduced various model structure of neural network for nonlinear

system control. Qian et al introduced the orthogonal network that applies distribution functions to transfer variables and to improve the problems of local minimum and slow convergence speed[10]. They also presented a method to minimize the Gibbs phenomenon in approximating piecewise continuous functions. The learning approach derived from the least square algorithm shows faster convergence speed than traditional backpropagation algorithm. In addition, multilayer neural network based on polynomial functions and function-link neural network with various transfer function were introduced in [11,12]. The latter was based on sin/cos functions and appeared to have faster convergence speed than traditional approaches. In [12], the researchers introduced a single-hidden-layer orthogonal neural network is developed by using orthogonal functions. Since the processing elements are orthogonal to one another and there is no local minimum of error function, the orthogonal neural network is able to avoid the local minimum problem. There are four well-known orthogonal function sets: Fourier series, Bessel functions, Legendre polynomials, Chebyshev polynomials. Among the four existing orthogonal functions, Legendre polynomials and Chebyshev polynomials have the properties of recursion and completeness. They are most suitable to generate the neural network. Some typical examples converge to show their performance in function approximation. The simulation results show that ONN has excellent convergence performance. Moreover, ONN is capable of approximating mathematic model of neural network.

In this paper, an orthogonal functions neural network is used to achieve the control of nonlinear systems. The adaptive controller is constructed by using Chebyshev orthogonal functions, which has advantages such as simple structure and fast convergence speed. The adaptive learning law of orthogonal neural network is derived to guarantee that the adaptive weight errors and tracking errors are bound by Lyapunov stability theory. Simulation results are given for a two-link robot in the end of the paper, and the control algorithm is validated.

2 Basic Theories of Orthogonal Functions and Orthogonal Neural Network

The orthogonal function neural network can approximate to any nonlinear function on the tight set, which has simple structure, fast convergence with the comparison of the common BP neural network. There are four orthogonal function sets: Fourier series, Bessel functions, Legendre polynomials, Chebyshev polynomials. Table 1 listed their properties related to the generation of ONN[12].

Table 1. The properties of the four orthogonal functions ^[12]

	Completeness at Boundary Points	Definition Interval	Recursive Property
Fourier series	May not exist	[0,T]	No
Bessel series	May not exist	[0,1]	Yes
Legendre series	Exist	0,1]	Yes
Chebyshev series	Exist	[0,1]	Yes

From Table 1, Legendre polynomials and Chebyshev polynomials are the two best choices to construct ONNs because they have recursive and completeness properties at the boundary points of their definition intervals. In the paper, the Chebyshev orthogonal polynomials are selected as the basis functions of the orthogonal function neural network. The ONN of Figure 1 is a typical single-output case. For a case with multiple outputs, its corresponding neural network will be composed of several single-output neural networks. The structure of an ONN with two outputs is constructed by two single-output ONNs. Since each of the two single-output neural networks has its independent weights, their respective weights can be trained separately.

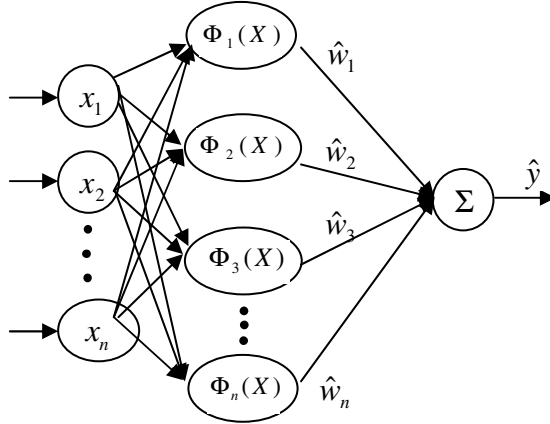


Fig. 1. An orthogonal neural network with two outputs

The Chebyshev polynomials are defined as the following form.

$$P_{j1}(x_j) = 1,$$

$$P_{j2}(x_j) = x_j,$$

$$P_{ji}(x_j) = 2x_j P_{j(i-1)}(x_j) - P_{j(i-2)}(x_j), \quad j = 1, 2, \dots, N, \quad i \geq 3. \quad \mathbf{x}_j \in [-1, 1] \quad (1)$$

According to the definition of ONN, the global output of the orthogonal function neural network is defined as Equation (2).

$$y = \sum_{i=1}^N W_i^T \Phi_i(x) \quad (2)$$

where $\Phi_i(x) = P_{i1}(x_1) \times P_{i2}(x_2) \times \dots \times P_{in}(x_n) = \prod_{j=1}^n P_{ji}(x_j)$, $P_{ji}(x_j)$ is Chebyshev polynomial.

The recursive relation among Chebyshev polynomials can not increase computation workload and complexity. For the example of ONN with the three inputs and the single output, the basis functions are settled as the following form.

$$\left\{ \begin{array}{l} \Phi_1(x) = 1 \\ \Phi_2(x) = x_1 x_2 x_3 \\ \Phi_3(x) = [2x_1 P_{12}(x_1) - P_{11}(x_1)][2x_2 P_{22}(x_2) - P_{21}(x_2)][2x_3 P_{32}(x_3) - P_{31}(x_3)] \\ \vdots \end{array} \right. \quad (3)$$

In addition, the training data will end up covering almost the entire interval $[-1,1]$ if time t keeps on increasing. A normalization of training samples $x_j(k)$ ($j=1,2,\dots,n$, $k=1,2,\dots$) will be necessary if the input domain is not in the interval $[-1,1]$. If the defined interval $[a,b]$ of training samples do not belong to $[-1,1]$, the variable x_j is transformed as

$$t_j = \frac{2}{b-a} x_j - \frac{b+a}{b-a}, \quad t_j \in [-1,1]$$

Lemma 1[11]. For any function $f(x)$ in the interval $[a,b]$ and any small positive number ε , $x \in R^n$, there exists an orthogonal function sequence.

$$\{\Phi_1(x), \Phi_2(x), \dots, \Phi_N(x)\}$$

The sequence $W_i (i=1,2,\dots,N)$ satisfies the following equation.

$$\left| f(x) - \sum_{i=1}^N W_i^T \Phi_i(x) \right| \leq \varepsilon \quad (4)$$

On the basis of Lemma 1, Lemma 2 is acquired as follows.

Lemma 2. For a given positive constant ε_0 and a continuous function $F(x) : x \in R^n$, exist an optimal weight vector $W = W^*$, $W^* = [W_1^*, W_2^*, \dots, W_N^*]^T$ to satisfy the following condition.

$$\|F(x) - W^{*T} \Phi(x)\| \leq \varepsilon_0 \quad (5)$$

where $\Phi(x)$ satisfies $\Phi(x) = [\Phi_1(x), \Phi_2(x), \dots, \Phi_N(x)]^T$.

3 The Dynamics Model of Robot Manipulator

In this paper, the model of the robot manipulator is built as a set of n rigid bodies connected in series with one end fixed to the ground and the other end free. The dynamic equations of robot manipulator motion are a set of highly nonlinear coupled differential equations. Using the Lagrange-Euler formulation, the dynamic equation of n -joint robot arm can be expressed as ^[1]

$$D(q)\ddot{q} + C(q, \dot{q})\dot{q} + G(q) + F_f(\dot{q}) = \tau - \tau_e \quad (6)$$

where q is the $n \times 1$ vector of joint angle. \dot{q} is the $n \times 1$ vector of joint angular velocity, and \ddot{q} is the $n \times 1$ vector of joint angular acceleration. $D(q)$ is the $n \times n$

matrix of symmetric position definite inertia. $C(q, \dot{q})$ is the $n \times 1$ vector of Coriolis and centrifugal. $G(q)$ is the $n \times 1$ vector of gravitational torques. $F_f(\dot{q})$ is the $n \times 1$ vector of dynamic and static friction forces. τ is the $n \times 1$ vector of joint torques supplied by the actuators, τ_e is the $n \times 1$ vector of compensating the unknown dynamics and external disturbances.

The controller of the robot manipulator includes a PD controller and orthogonal function neural network (ONN). The ONN controller is connected in parallel with the PD controller to generate a compensated control signal. The control law is given as the following form.

$$\tau = \tau_{ONN} + \tau_{PD} \tag{7}$$

where τ_{ONN} is the output torque of the ONN, and τ_{PD} is the output torque of PD controller, satisfying $\tau_{PD} = EK_{PD}$, $K_{PD} = \begin{bmatrix} k_p & k_d \end{bmatrix}$. The tracking error vector is defined as

$$E = [q_d - q \quad \dot{q}_d - \dot{q}] = [e \quad \dot{e}]^T \tag{8}$$

where the variable q_d is the desired joint angle, and e is the tracking error. If the parameters of robot dynamic model are known, the control torque can be designed as

$$\tau^* = D(q)\ddot{q}_d + C(q, \dot{q}) + G(q) + F_f(\dot{q}) + \tau_e + D(q)KE \tag{9}$$

where K is $K = \begin{bmatrix} k_2 & k_1 \end{bmatrix}$, and k_2, k_1 are positive real numbers. Substituting (9) into (6) yields

$$\ddot{e} + k_1\dot{e} + k_2e = 0 \tag{10}$$

If the proper K is chosen, the tracking error will converge to zero. However, the external disturbances and uncertainties are unknown in practice. We proposed the orthogonal neural network as the torque controller, and the perfect control law is executed by

$$\tau^* = \tau_{ONN}^* + D(q)KE \tag{11}$$

4 The Stability Analysis of Orthogonal Neural Network Controller

From (6) to (9), the error tracking equation is arranged as the following form.

$$\dot{E} = AE + B(\tau^* - \tau_{ONN} - \tau_{PD}) \tag{12}$$

where A satisfies $A = -\begin{bmatrix} 0 & -I \\ k_2I & k_1I \end{bmatrix}$, $B = \begin{bmatrix} 0 \\ D(q)^{-1} \end{bmatrix}$.

Substituting (11) and (7) into (12), we have

$$\begin{aligned}\dot{E} &= AE + B \left[\tau_{ONN}^* - \tau_{ONN} + D(q)KE - K_{PD} \right] \\ &= \hat{A}E + B \left[(W^*)^T \Phi - W^T \Phi \right]\end{aligned}\quad (13)$$

where \hat{A} satisfies $\hat{A} = - \begin{bmatrix} 0 & -I \\ D(q)^{-1}k_p & D(q)^{-1}k_d \end{bmatrix}$, W^* and Φ are the optimal weights and Chebyshev orthogonal basis functions, respectively. K_{PD} is the PD controller gain, $K_{PD} = \begin{bmatrix} k_p I & k_d I \end{bmatrix}$. Equation (13) can be written as

$$\dot{E} = \hat{A}E + B \left[(W^* - W)^T \Phi \right] \quad (14)$$

Theorem. Exist a positive definite and symmetric matrix P , and satisfy $P\hat{A} + \hat{A}^T P = Q$, where Q is a positive definite and symmetric matrix, the controller is designed as

$$\dot{W} = k\Phi E^T P B \quad (15)$$

where W is bounded, and define the constraint set Γ for W as $\Gamma = \{\|W\| \leq \|W_0\|\}$.

Proof. The Lyapunov function is defined as the following form.

$$V(t) = 0.5k^{-1}tr \left[(W^* - W)^T (W^* - W) \right] + 0.5E^T P E \quad (16)$$

Differentiating equation (16), using (14) and $P\hat{A} + \hat{A}^T P = Q$, we have

$$\begin{aligned}\dot{V}(t) &= 0.5\dot{E}^T P E + 0.5E^T P \dot{E} - k^{-1}tr \left[(W^* - W) \dot{W} \right] \\ &= -0.5E^T Q E + E^T P B (W^* - W)^T \Phi \\ &\quad - k^{-1}tr \left[(W^* - W) \dot{W} \right]\end{aligned}\quad (17)$$

Under the condition (15), (17) becomes the following equation.

$$\begin{aligned}\dot{V}(t) &= -0.5E^T Q E + E^T P B (W^* - W)^T \Phi - tr \left[(W^* - W) \Phi E^T P B \right] \\ &= -0.5E^T Q E + E^T P B (W^* - W)^T \Phi - tr \left[E^T P B (W^* - W) \Phi \right] \\ &= -0.5E^T Q E \leq 0\end{aligned}\quad (18)$$

If and only if $E = 0, V(t) = 0$. Therefore, the global stability is guaranteed by the Lyapunov theorem.

Remark 1. When conventional NN are used for control purpose, their structures are difficult to determine. In order to guarantee the approximation accuracy, more layers

and more neurons should be used. Hence, the well-known “explosion of terms” phenomenon occurs and the convergent speed is greatly decreased. As a result, such NNs are not suitable for real time control. Due to the orthogonal basis functions, the presented ONN has both a simple structure and a relatively fast convergent speed.

Remark 2. With the ONN controller, only measurable and local information is used and there is no requirement for the system’s model. The system is handled as a grey box.

Remark 3. It is well known that when we use an adaptive control algorithm, the estimated parameters will not always converge to their true values although the stability of the system is guaranteed [13]. Derived from the Lyapunov theory, the proposed ONN learning algorithm makes it possible to drive the weights to their optimal values and therefore a global minimum is achieved. In case that the given task is repeatable, the final weights in one trial can be used as the defined values for the each trial.

5 Simulation

Simulations were carried out to verify that the proposed ONN could compensate for uncertainties disturbances. The manipulator used for the simulation study is a typical two degree-of-freedom robot. The dynamic equation of the manipulator and the parameters were taken from[1].

$$D(q)\ddot{q} + C(q, \dot{q})\dot{q} + G(q) + F_f(\dot{q}) = \tau - \tau_e$$

where

$$D(q) = \begin{bmatrix} 2.8 + 2 \cos q_2 & 0.7 + \cos q_2 \\ 0.7 + \cos q_2 & 0.9 \end{bmatrix}, \quad C(q) = \begin{bmatrix} -\dot{q}_2 \sin q_2 & -(\dot{q}_1 + \dot{q}_2) \sin q_2 \\ \dot{q}_1 \sin q_2 & 0 \end{bmatrix}$$

$$G(q) = 0, \quad F_f(\dot{q}) = \text{diag}[2 \text{sgn}(\dot{q}), 2 \text{sgn}(\dot{q})], \quad k_p = \text{diag}[15, 15], \quad K_d = \text{diag}[8, 8],$$

$$q_d = [0.5 \cos t + 0.2 \sin 3t, -0.2 \sin t - 0.5 \cos t]^T$$

where q_d is the desired trajectory. The number of the input layer of the orthogonal neural network is 3 .The number of the hidden layer is 15, $N = 15$. The simulation results are shown as follows.

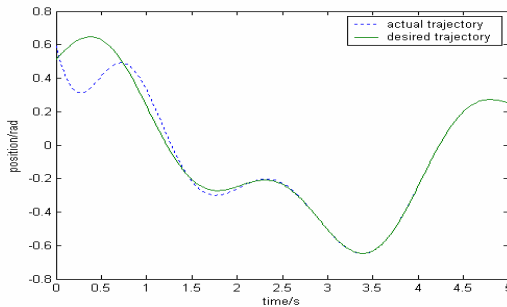


Fig. 2. Tracking of q_1 using PD control arithmetic

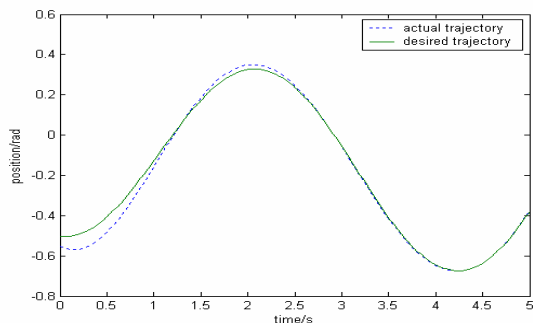


Fig. 3. Tracking of q_2 using PD control arithmetic

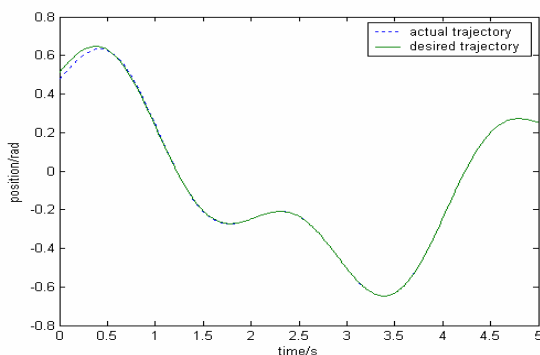


Fig. 4. Tracking of q_1 using the proposed method

With the comparison between the PD controller and the proposed scheme, the performance of the proposed controller is superior to that of the PD controller.

In Table 2, we compare our control method with other control methods by using same model. It can be seen that the performance of our method is superior to that of other control methods.

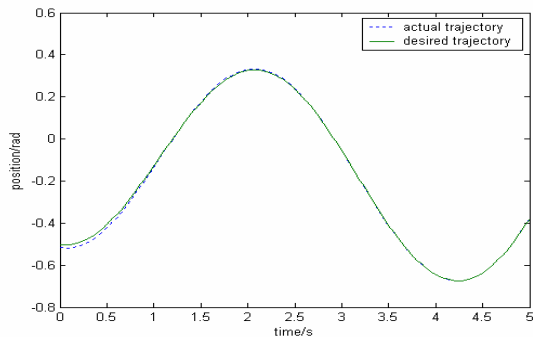


Fig. 5. Tracking of q_2 using the proposed method

Table 2. Generalization Result Comparisons

Methods	Training Cases	Track Mean Square Errors
PD control	Tracking of q_1	0.068
	Tracking of q_2	0.035
Our Method	Tracking of q_1	0.0034
	Tracking of q_2	0.0015

6 Conclusion

The controller of robot manipulators is designed by these conventional feedback controllers, such as PID controllers. However, when these conventional feedback controllers are directly applied to nonlinear systems, they suffer from the poor performance and low robustness due to the uncertainties and the external disturbances. In this paper, the adaptive controller is constructed by using Chebyshev orthogonal functions neural network. The adaptive learning algorithm of orthogonal neural network is derived to guarantee that the adaptive weight errors and tracking errors are bound by using Lyapunov stability theory. Simulation results are given for a two-link robot in the end of the paper, and the control scheme is validated.

References

1. Seraji, H.: Approach to Multivariable Control of Manipulators. *Journal of Dynamic Systems, Measurement and Control*, Transactions ASME 109, 146–154 (1987)
2. Seraji, H.Y.: Decentralized Adaptive Control of Manipulators: Theory, Simulation, and Experimentation. *IEEE Transactions on Robotics and Automation* 5, 183–201 (1989)
3. Li, Y.M., Ho, Y.K., Chua, C.S.: Model-Based PID Control of Constrained Robot in a Dynamic Environment with Uncertainty. In: *IEEE Conference on Control Applications – Proceedings: ICA*, Anchorage, USA (2000)
4. Wang, K.J., Tang, M., Liu, W.: A Study of Fuzzy Chaotic Neuron and Fuzzy Chaotic Neural Network. *IEEE International Conference on Mechatronics and Automation: IMA*, Niagara Falls, ON, Canada (2005)
5. Karimi, H.R., Babazadeh, A.: Modeling and Output Tracking of Transverse Flux Permanent Magnet Machines Using High Gain Observer and RBF Neural Network. *ISA Transactions*, 445–456 (2005)
6. Ciftcioqlu, O., Sariyildiz, I.S.: Enhanced Multivariable TS Fuzzy Modeling in Neural Network Perspective. In: *Annual Conference of the North American Fuzzy Information Processing Society*, pp. 150–155 (2005)
7. Wang, K.J., Tang, M., Liu, W.: A Study of Fuzzy Chaotic Neuron and Fuzzy Chaotic Neural Network. In: *IEEE International Conference on Mechatronics and Automation*, pp. 890–895. IEEE Press, New York (2005)
8. Kryzhanovsky, B., Magomedov, B.: Application of Domain Neural Network to Optimization Tasks. In: Nagel, W.E., Walter, W.V., Lehner, W. (eds.) *Euro-Par. LNCS*, vol. 3628, pp. 397–403. Springer, Heidelberg (2005)

9. Ye, B., Guo, C.X., Cao, Y.J.: Identification of Fuzzy Model Using Evolutionary Programming and Least Square Estimate. In: IEEE the International conference of Fuzzy System, Budapest, Hungary, Huly, New York, pp. 25–29 (2004)
10. Qian, S., Lee, Y.C., Jone, R.D., Barnes, C.W.: Function Approximation with an Orthogonal Basis Net. In: International Joint Conference Neural Networks: IJNN III, Detroit, MI, United States (1992)
11. Pao, Y.H., Phillips, S.M., Sobajic, D.J.: Neural-Net Computing and the Intelligent Control of Systems. *International J. of Control.*, 263–289 (1992)
12. Chen, F.S., Ching, S.T., Chen, S.C.: Properties and Performance of Orthogonal Neural Network in Function Approximation. *International Journal of intelligent systems* 16, 1377–1392 (2001)
13. Spong, M.W., Vidyasagar, M.: *Robot Dynamics and Control*. John Wiley & Sons, Chichester (1989)

Q-Learning Based on Dynamical Structure Neural Network for Robot Navigation in Unknown Environment

Junfei Qiao, Ruiyuan Fan, Honggui Han, and Xiaogang Ruan

Institute of Intelligence System, College of Electronic Information and control Engineering,
Beijing University of Technology, Beijing 100000, China
adqiao@sina.com, fanruiyuan@gmail.com, rechard112@163.com

Abstract. An automation learning and navigation strategy based on dynamical structure neural network and reinforcement learning was proposed in this paper. The neural network can adjust its structure according to the complexity of the working environment. New nodes or even new hidden-layers can be inserted or deleted during the training process. In such a way, the mapping relations between environment states and responding action were established, and the dimension explosion problem was solved at the same time. Simulation and Pioneer3-DX mobile robot navigation experiments were done to test the proposed algorithm. Results show that the robot can learn the correct action and finish the navigation task without people's guidance, and the performance was better than artificial potential field method.

Keywords: Mobile robot, Navigation, Reinforcement learning, Dynamical neural network.

1 Introduction

Robot navigation is a complex system including environment sense, dynamical decision making, and actions control and so forth. As its application expansion in aerospace, medical services, industrial production and many other important fields, mobile robot navigation attracts more and more researchers[1, 2]. Often the robot working environment is unpredictable and volatile, so the robot is expected to learn the environment and be able to make decision himself. Q-learning has a character that independent of the environment model and learning on-line, So Q-learning is considered as a promising machine learning strategy, especially in the unknown environment navigation[3].

The working environment and action space should be separated in classical Q-learning, but this division leads to dimension explosion problem[4]. To overcome the defect in classical Q-learning, some methods were proposed such as bind neural network and Q-learning together[5,6,7,8]. Of course, this is an effective method, but the structures of these networks were fixed and the information processing capacity was limited. In this paper, a neural network model called DSNN (Dynamical Structure Neural Network) was proposed. The network has a flexible structure and new hidden nodes or even new hidden layers can be inserted while redundant nodes and layers can

be deleted. Based on such a strategy, the size of network is match to the application problem. The proposed method was applied on Pioneer3-DX mobile navigation in corridor environment, results show that this method is effective and the performance is better than APF (Artificial Potential Field) method.

2 Mobile Robot Navigation Architecture

Pioneer3-DX mobile robot was used in the experiment, 16 sonar sensors were equipped in this robot and the sense range is 0~5000mm. Also two wheels were equipped, at the rear of the robot, there is a small wheel which can make the robot rotate to any direction. The Pioneer3-DX and its sonar sensors were shown in Fig.1 and Fig.2.



Fig. 1. Pioneer3-DX mobile robot

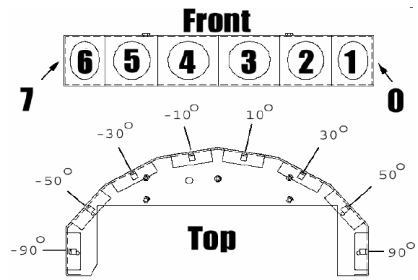


Fig. 2. Front sonar of the robot

As showed in fig.3, this is the mobile robot navigation diagram. The sonar array scanning environment and detect obstacles around the robot. At the same time, the robot position is available through Odometer. Here, we define a vector $S = \{d_1, d_2, d_3, d_4, d_5\}$ as the description of robot working environment state space, and define $A = \{a_1, a_2, a_3, a_4, a_5\}$ as the robot action space. A neural network called DSNN (Dynamical Structure Neural Network) was built to replace Q-table in classical reinforcement Q-learning. There are two reasons to use DSNN, firstly, the neural network can avoid the dimension explosion problem which was a key problem in Q-learning; secondly, this network has a flexible structure and can adjust the size of nodes and weights according to the application case. The robot has no prior knowledge about the working states and the mission to accomplish at the beginning. A “reasonable” action was selected from the action space every training circle, and then the robot working states were changed. The executed action is rewarded if the current state is better than the prior one, for example the robot is closer to the goal or farther to the obstacles or both, on the contrary, this action is punished. After many times trainings, the robot begins to know what should do in current state and what should not. Two wheels were controlled by two independent motors and the robot can move forward or back or rotation to any direction.

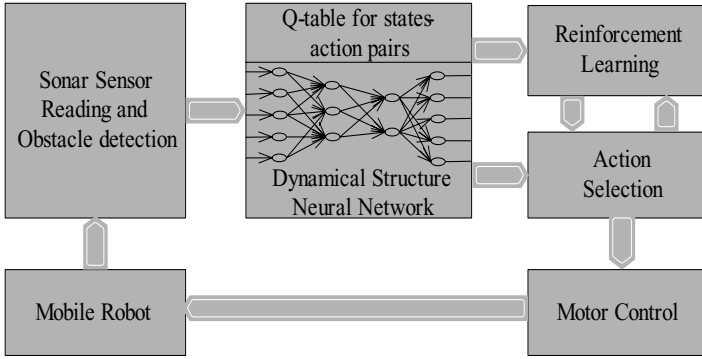


Fig. 3. Robot navigation system

3 Navigation Based on Reinforcement Learning and DSNN

3.1 Reinforcement Q-Learning

As mentioned before, we had defined state space $S = \{d_1, d_2, d_3, d_4, d_5\}$ and action space $A = \{a_1, a_2, a_3, a_4, a_5\}$

Define $d = \min(d_1, d_2, d_3)$ as the nearest distance from robot to obstacles.

a. The robot was expected to approach the goal

$$\begin{cases} r_1(t) = d_4(t+1) - d_4(t) \\ r_1(t) < 0 \end{cases} \quad (1)$$

b. The robot is expected to avoid obstacles

$$\begin{cases} r_2(t) = d(t+1) - d(t) \\ r_2(t) > 0 \end{cases} \quad (2)$$

c. The robot turn to the goal direction

$$\begin{cases} r_3(t) = d_5(t+1) - d_5(t) \\ r_3(t) < 0 \end{cases} \quad (3)$$

We expect the robot to approach the goal, avoid obstacles and turn to the goal direction, so the reward can be written as:

$$r(t) = -\alpha r_1(t) + \beta r_2(t) - \gamma r_3(t) \quad (4)$$

Where α, β, γ are some factors and $t=1, 2, 3, 4, 5, \dots$

Table 1. State Space and Action Space in Reinforcement Learning

State Space	Description	Action Space	Description
d_1	Distance between robot and left obstacles.	a_1	Rotate +15°and move 100mm
d_2	Distance between robot and front obstacles.	a_2	Rotate -15°and move 100mm
d_3	Distance between robot and right obstacles	a_3	Rotate +10°and move 100mm
d_4	Distance between robot and goal	a_4	Rotate -10°and move 100mm
d_5	Angle between robot moving direction and goal	a_5	Rotate 0°and move 100mm

The aim of Q-learning is to learn the state-action pair value $Q(s, a)$ which is the maximum discounted amount of reward. Thus Q-function can be written as:

$$Q_t(s_t, a_t) \leftarrow r(t) + \rho \max_{a_k \in A} Q_{t-1}(s_{t+1}, a_k) \tag{5}$$

$$\Delta Q_t(s_t, a_t) \leftarrow r(t) + \rho \max_{a_k \in A} Q_{t-1}(s_{t+1}, a_k) - Q_{t-1}(s_t, a_k) \tag{6}$$

As its name, reinforcement learning is a learning process, in the beginning, the left part and right part of (5) is not equal, out destination is to make the $\Delta Q_t(s_t, a_t)$ become smaller and smaller, that means the reinforcement is convergence and the training process is to be finished. In this paper, DSNN was used to approximate the relationship between states space and action space in the Q-learning.

3.2 Dynamical Structure Neural Network

In order to avoid the problem of dimension explosion and make the network to adjust its structure himself, we established a flexible structure neural network. In the

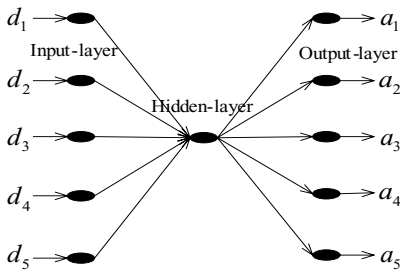


Fig. 4. Initialed structure of DSNN

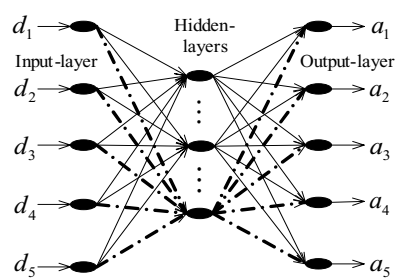


Fig. 5. Inserting new node to hidden layer

beginning, the network was initialed as a three-layered network which includes five nodes as input layer and another five nodes as output layer. The initialed network just has one hidden layer and only one hidden node in this layer, shown in fig.4. The five input-layer nodes were corresponding with the state space S while the five output-layer nodes were related to the action space A . The number of input-layer nodes and output-layer nodes were fixed and do not change in the training process, while the hidden layers and number of nodes in the hidden layer was tuned.

We define $E(l, n)$ as the current training error, where l is the number of hidden-layers while n is the number of neural nodes in the hidden-layer which adjacent to output-layer.

$$E(l, n) = \frac{1}{2} \sum_{p=t-\lambda}^t \sum_{k=1}^5 (\Delta Q_p(s_p, a_k))^2 \tag{7}$$

Where λ is an integer and means we accumulate error for λ times then get a $E(l, n)$. In such a way, there was a “window” and the training process is rolled.

The net was begun with a simple structure, and it is possible that the network is not complex enough to have the capacity to learning the relationship between state-action pairs. If enough training has been carried out while the training error deduction was slim, new node should insert and further training is needed. Described as fig.5 and equation (8), where p is a small integer.

$$\begin{cases} |E(l, n) - E(l, n - p)| / E(l, n) > \xi \\ E(l, n) > E_0 \end{cases} \tag{8}$$

There are often some cases that nodes were inserted in the same layer more and more but the training error reduction is not Significant. Now we should consider adding a new hidden-layer. Refer to fig.6 and equation (9). Experiments and test show that mort closer to the output-layer the weights tuned much, so it is easy to comprehend that inserting new hidden-layer adjacent to output-layer was reasonable.

$$\begin{cases} |E(l, n) - E(l, n - p)| / E(l, n) \leq \xi \\ E(l, n) > E_0 \end{cases} \tag{9}$$

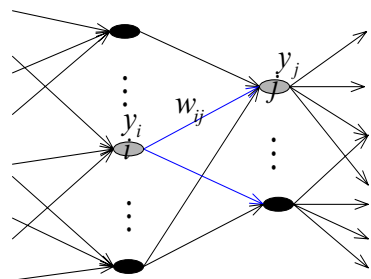
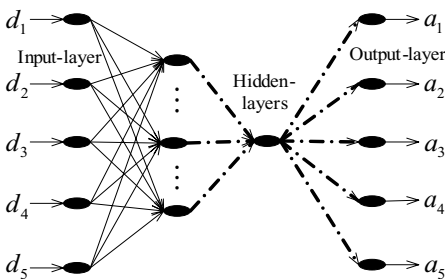


Fig. 6. Inserting new hidden-layer to the network Fig. 7. Imp of neuron nodes in the network

After every inserting, no matter new node or new hidden-layer, the net should re-train for enough times until the error reduction is small. There maybe some redundant nodes in the network when the error threshold had reached. These nodes have little positive effect to the network but may lead to bad generalization ability. So it is necessary to delete these nodes. Joog-Sock LEE proposed a deconstruction method called IMF (Impact Factor)[9]. The key idea of this method was to evaluate every node's effect to the next layer. Shown as fig.7, this is a part of the network, input of the neural j in the next layer for the m -th training input x^m , can be written as following:

$$x_j^m = \sum_i w_{ji} y_i^m + b_j \tag{10}$$

$$x_j^m = \sum_i w_{ji} (y_i^m - \bar{y}_i) + \sum_i w_{ji} \bar{y}_i + b_j \tag{11}$$

Where w_j^m is the input of node j in the next layer, \bar{y}_i is the average output value of i -th neuron for all training data. The total amount of contribution of the i -th neuron to the next layer can be defined as $\sum_j w_{ji}^2 (y_i^m - \bar{y}_i)$, and the i -th neuron's Imf can be written as:

$$Imf_i = \sum_j w_{ji}^2 \sigma_i^2 \tag{12}$$

To a neural node which has a small imf had a small contribute to the network and waste the training time, more worse, these redundant neurons lead to a bad generalization ability.

3.3 Q-Learning Based on Dynamical Structure Neural Network

Robot sense environment through sonar and Odometer and the environment state information was imported to DSNN, and then a action was selected from the action space and carried out. Such the working states changed and the robot get a reward or punishment. The weights and bias of the DSNN was tuned and network structure is adjusted until the error threshold was reached. In the early stages of learning, the main task is to explore environment, so the select randomness should bigger, On the contrary, in the latter stages of learning, the main task is to make the training converge, the select randomness should be smaller. So Boltzmann Annealing algorithm was used.

$$P(a_k) = \frac{e^{Q_i(s_t, a_k)/T}}{\sum_{a_k \in A} e^{Q_i(s_t, a_k)/T}} \tag{13}$$

Where T is the virtual temperature. $T = T_0 t^{-1/\tau}$, the training procedure is as below:

Step1: Initial the network and other parameters.

Step2: Obtain the current environment states and import to DSNN, compute the net output.

Step3: Compute $p(a_k)$ and select a action in roulette gambling principle.

Step4: Carry action a_k and reading new states, then compute the reward $r(t)$ and error

$$E(l, n)$$

Step5: Adjust the weights and structure of the network.

Step6: If $E(l, n) < \delta$ turn to step7, else turn to step2. Where δ is the error threshold.

Step7: Save the network structure and weights, and then end the learning process.

4 Results of Experiment and Analysis

4.1 Comparative Analysis of Simulations

In the Pioneer3-DX robot platform MobileSim we created the mobile robot working environment and compared three robot navigation methods, the first one is reinforcement learning based on DSNN and the second is navigation based on APF (Artificial Potential Field) method, and the third is reinforcement learning based on BPNN(Back-Propagation Neural Network)

It was seen clearly that there was a wall between the start point and goal point in the three figures. When the reinforcement learning based on DSNN is used, after training or learning process, the robot can avoid the wall and reach the goal point, showed in fig.8. But to artificial potential field method, the robot was trapped in the corner of wall and can't escape. Because there maybe some balance points in the potential field, once the robot arrived near these points, often the robot was attracted to the balance point and trapped, showed in fig.9. When the reinforcement learning method based on BPNN is used, the robot also can finish the navigation mission, show as fig.10, but the hidden layer nodes were defined as 15 before training process. The DSNN has an obvious advantage is that the number of hidden nodes can be adjusted according to specific issues. To illustrate this point, experiments were done for ten times, fig.11 described the numbers of hidden nodes of each time. The number of each time was not equal, because the action sequence was different every time, but it should be noted that the number of neurons was relatively stable in a long time and there was a small difference.

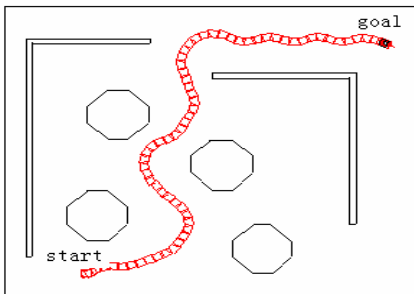


Fig.8. Simulation trajectory baaed on DSNN

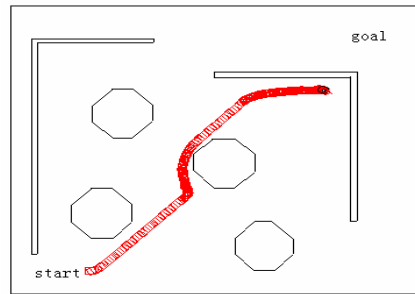


Fig.9. Simulation trajectory based on APF

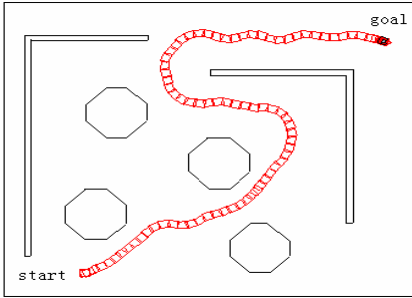


Fig.10. Simulation trajectory based on BPNN

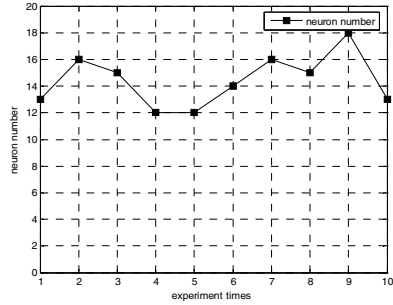


Fig.11. Hidden nodes number of DSNN

4.2 Comparative Analysis of Pioneer3-DX Navigation Experiment

In order to verify the validity of the algorithm, Pioneer3-DX mobile robot navigation experiments were executed. The working environment was the laboratory corridor, the width of the corridor is 1740mm, the most narrow place is 1540mm, show as fig.13 and fig.14, the coordinate of starting point is Start(6000mm,870mm) and the coordinate of goal point is Goal(32030mm, 9345mm).



Fig.12. Pioneer3-DX robot navigation in corridor environment

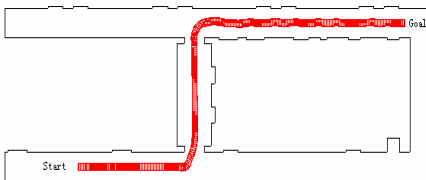


Fig.13. Navigation result based on DSNN

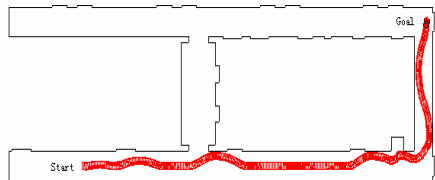


Fig.14. Navigation result based on APF

Compare the two figures, the trajectory of DSNN was nearly in the middle of the corridor and is smooth while the navigation trajectory based on artificial potential field method is not as good, many times, the robot knocked against the walls, and the robot must be paused and restart again. The simulation result is consistent with the experimental one. Though the navigation based on DSNN need much time to learn the environment and train the network. Robot navigation experiment based on BPNN was also done, the result was similar to DSNN, but you should define the number of hidden layer first, in the experiment, a BPNN was created with a structure of 5-15-5.

5 Conclusion

Mobile robot autonomous navigation is a common interest to many artificial intelligence researchers. Use of reinforcement learning strategy to achieve mobile robot navigation is an effective way, Combination of Q-learning and dynamical structure neural network not only solve the problem of dimension explosion but also create a structure self-organizing strategy, the network can adjust the structure and weights. To a certain extent, the robot became clever enough to contact with the environment and learn himself. Simulation and experiment were done and results show the effect of Q-learning based on DSNN get a better navigation perform than artificial potential field method.

Acknowledgements. This work is partially supported by National Science Foundation of China(No.60674066,60873043), the National High Technology Research and Development Program(“863” Program) of China (No.20061D0501500203).

References

1. Wollherr, D., Buss, M.: Human-robot Collaboration: A Survey. *International Journal of Humanoid Robotics* 5, 47–66 (2008)
2. Jan, G.E., Chang, K.Y., Parberry, I.: Optimal Path Planning for Mobile Robot Navigation. *IEEE-ASME Trans. On Mechatronics* 13, 451–460 (2008)
3. Lucian, B., Robert, B., Schutter, B.D.: A Comprehensive Survey of Multiagent Reinforcement Learning. *IEEE Trans. On Systems, Man, and Cybernetics* 38, 156–172 (2008)
4. Carrersa, M., Yub, J., Batlle, J., Ridao, P.: Application of SONQL for Real-time Learning of Robot Behaviors. *Robotics and Autonomous System* 55, 628–642 (2007)
5. Arleo, A., Smeraldi, F., Gerstner, W.: Cognitive Navigation Based on Bonuniform Gabor Space Sampling Unsupervised Growing Networks and Reinforcement Learning. *IEEE Trans. On Neural Networks* 15, 639–652 (2004)
6. Kumar, S.: *Neural Network: A Classroom Approach*, International edition. McGraw-Hill Publishing Company Limited, New York (2005)
7. Ma, X.L., Konstantin, K.L.: Global Reinforcement Learning in Neural Networks. *IEEE Trans. On Neural Networks* 18, 573–577 (2007)
8. Tan, A.H., Lu, N., Xiao, D.: Integrating Temporal Difference Methods and Self-Organizing Neural Networks for Reinforcement Learning with Delayed Evaluative Feedback. *IEEE Trans. On Neural Networks* 19, 230–244 (2008)
9. Lee, J.S., Lee, H., Kim, J.Y., Nam, D.Y., Park, C.H.: Self-Organizing Neural Networks by Construction and Pruning. *IEICE Trans. Inf. & Sys. E* 87, 2489–2498 (2004)

Research on Mobile Robot's Motion Control and Path Planning^{*}

Shigang Cui¹, Xuelian Xu², Li Zhao¹, Liguo Tian¹, and Genghuang Yang¹

¹School of Automation and Electric Engineering,
Tianjin University of Technology and Education, 300222, Tianjin, China
{Shigang Cui, Li Zhao, Liguo Tian, Genghuang Yang}cuisg@163.com

²School of Mechanical Engineering, Tianjin University of Technology and Education,
300222, Tianjin, China
Xuelian Xu hehe2006_9@163.com

Abstract. In this article, the kinematics modeling for the practical robot and movement formula have been established, whose are considered by the feasibility and reliability of actual control, and controls the movement of mobile robot by the planning route, guides the robot to complete the mission. On the other hand, ant colony optimization algorithm is used to solve the robot path planning. Based on the original ant colony optimization algorithm, it modifies the route choice strategy and the pheromone updating strategy etc. according to the information provided by the global map and the mission to complete. And the experimental results on MATLAB are convinced the optimal path.

Keyword: mobile robot, kinematics modeling, motion control, path planning, ant colony algorithm.

1 Introduction

The path planning is a very important branch of mobile robots' research, and plays a vital role of robot's navigation, and is also a significant manifestation of the intelligence level of mobile robot [1]. The kinematics is the most basic research on how the robot hardware system to move, and also is the foundation of realizing the path planning algorithms. In this article, the kinematics model for the actual robot have been established which is the mathematic base for motion control.

It has developed many methods in path planning field for many years, however those algorithms are some insufficiency in some degree [2], [3], [4]. The ant colony optimization (ACO) algorithm [5] simulates and modifies the behavior of natural ants searching for food. This article is based on this algorithm, and modified it according to the actual requests, obtained the appropriate path, and proved it through the experiments.

^{*} This work was supported by National High-tech R&D Program (863 Program), 2007AA04Z254, Tianjin Binhai New Area's Construction Science and Technology Action Planning Project Supported by Chinese Academy of Sciences, TJZX2-YW-06, and the key project of Tianjin Science and Technology Planning, 08ZCKFSF03400.

2 The Kinematics Modeling and Control for TUT06_B

The robot which is called TUT06_B [6] uses the twin pedrail type movement structure. After power on, motors turning, they driver the coaxial planetary gear reducer and the input stage gear of the right angle reducer through the coupling, and then drive the driving pulley to revolve through the output stage gear of the right angle reducer, thus lead the pedrail to move. Using this structure, it can improve to the robot's movement stability and climbing performance effectively. Moreover, it's equipped two assistant wheels in the side of the pedrail. As shown in Fig.1, it's the motion part of the mobile robot.

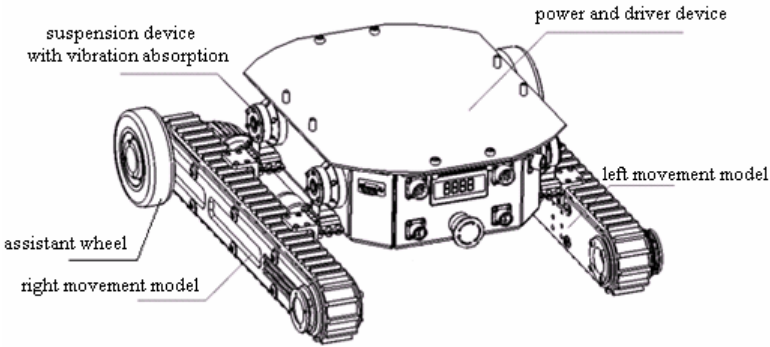


Fig. 1. Motion part of the mobile robot

2.1 Position indication of TUT06-B [1] [7]

Although this robot's motion part has used the pedrail structure, it doesn't have any differences with the wheeled robot in principle. In this article, it is simplified to twin wheel differential driving structure, and the robot is defined a rigid body on the driving pulley, which is moving in the horizontal plane, as Fig.2 (a) showing. Establishes the following coordinate relations: O-XY is the workspace's global reference coordinate, and $O_R-X_R Y_R$ is the mobile robot's local reference coordinate, O_R is the origin of $O_R-X_R Y_R$, which is superposition with the right driving pulley's middle point. O_R 's position in O-XY is (x_{OR}, y_{OR}) , and the angle difference between two coordinate is Φ , thus the robot's pose in O-XY is $\xi = (x_{OR}, y_{OR}, \phi)^T$.

In order to analyze the robot's movement, it should map the point's position in global reference coordinate to the robot local reference coordinate, which satisfied:

$$\xi_R = R(\phi) \cdot \xi \tag{1}$$

$$R(\phi) = \begin{bmatrix} \cos \phi & \sin \phi & 0 \\ -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

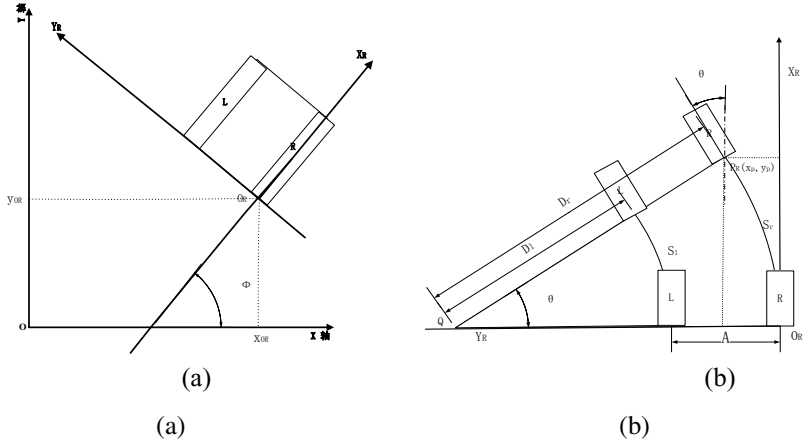


Fig.2. Kinematics modeling of the mobile robot

2.2 Modeling for Direct Kinematic Problem

The so-call direct kinematic problem is to obtain the robot's movement route with known the two wheel's speeds. As shown in Fig.2 (b), supposed the robot is at the initial point, which means the right wheel's position namely is origin of O-XY. Let the distance between the two pedrail is A, the driving pulley's radius is r, and the left and right driving pulley's angular velocity is $\dot{\omega}_l$ and $\dot{\omega}_r$, respectively. According to formula 1, the robot's movement coordinates in local reference coordinate maps to global reference coordinate, which satisfied:

$$\dot{\xi} = R(\phi)^{-1} \cdot \dot{\xi}_R \tag{3}$$

$$R(\phi)^{-1} = \begin{bmatrix} \cos \phi & -\sin \phi & 0 \\ \sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4}$$

Because each wheel's relative position of the robot is fixed, the wheel isn't able to move along the Y_R direction, namely $y_R=0$. If the right wheel is rotating and the left wheel is stopping, the rotation angle θ_r is satisfied $\theta_r = \frac{2\pi r \cdot \dot{\omega}_r}{A}$. In the same way,

$$\theta_l = -\frac{2\pi r \cdot \dot{\omega}_l}{A}$$

Therefore, the kinematics modeling is as follows:

$$\dot{\xi} = R(\phi)^{-1} \cdot \dot{\xi}_R = \begin{bmatrix} \cos \phi & -\sin \phi & 0 \\ \sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \frac{2\pi r \cdot \dot{\omega}_r + 2\pi r \cdot \dot{\omega}_l}{A} \\ 0 \\ \frac{2\pi r \cdot \dot{\omega}_r}{A} - \frac{2\pi r \cdot \dot{\omega}_l}{A} \end{bmatrix} \tag{5}$$

2.3 Modeling for Inverse Kinematics Problem

The so-call inverse kinematic problem is to obtain the motion law of the robot’s driving wheels with known the planned path, which means that it should obtain $\dot{\omega}_i$ and $\dot{\omega}_r$ with known $\dot{\xi} = (x_R, y_R, \theta)^T$. The so-called mobile robot's motion control is to guide the robot moving along the expectation path by adjusting the robot's speed and direction.

In order to simplify the calculation, it is analyzed in local reference coordinate, as shown in Fig.2 (b). At time t, the robot has rotated an arc with central angle θ (anti-clockwise is defined positive direction), and arrives at position P_R , the left and right driving pulley have passed distance S_l and S_r respectively.

$$S_l = 2\pi r \omega_l t = D_l \cdot \theta = (D_r - A) \cdot \theta \tag{6}$$

$$S_r = 2\pi r \omega_r t = D_r \cdot \theta \tag{7}$$

According to the relation between arc and chord, it can obtain formula as follows

$$2D_r \cdot \sin \frac{\theta}{2} = \sqrt{x_p^2 + y_p^2} \tag{8}$$

And angular velocity can obtain from formula (7) and (8) simultaneous.

$$\omega_r = \frac{\theta \cdot \sqrt{x_p^2 + y_p^2}}{4\pi r t \sin \frac{\theta}{2}} \tag{9}$$

$$\omega_l = \frac{\theta \cdot (\sqrt{x_p^2 + y_p^2} - 2A \sin \frac{\theta}{2})}{4\pi r t \sin \frac{\theta}{2}} \tag{10}$$

According to formula (9) and (10), when the robot is hoped to move along a straight line, which means $\theta=0$, its angular velocity should set to $\omega_l = \omega_r$; when the robot is hoped to turn left, namely $\theta>0$, its angular velocity should set to $\omega_l < \omega_r$; and vice versa. Based on the kinematics inverse problem's modeling, if the planned route is determined, namely the next position is arranged, and the robot could arrive at the target point at the set speed.

From the above analysis, it is known that the robot’s turning is realized by setting different angular speed for wheels [8], especially, when $\omega_l = -\omega_r$, namely the two wheel are rotating at the same speed but in the opposite direction, the robot will rotate at the original-place. In this way, robot just rotates a certain angle without displacement in forward or back direction when turning. Therefore, mobile robot’s movement can divide to three kinds: linear motion, arc motion and rotational motion. Any sub-paths can be made of these three kinds movement. In the experiment, for reducing calculation works, the arc motion is simplified to the union of linear motion and rotation motion, which means the arc is carried on the fitting with certain line segments.

Thus, the method to arrive the expected position is to revolve certain angle which is decided by the angle between the robot's current direction and expected direction, and then to move along the line. In the next part of this article, the path planning algorithm is modified to adapt to the actual motion control, in this way, it can realize the algorithm more easily, and also can improve real-time performance of system.

3 The Modified ACO Algorithm Using in Robot's Path Planning

The first application of ACO is travelling salesman problem (TSP) [5]. Therefore the mathematical model is built according to TSP[9][10]: n is the total of cities, r is the set of cities, $\Gamma = \{r_1, r_2, \dots, r_n\}$, m is the total of ants, $b_i(t)$ expresses the number of ant located in the city i at time t , d_{ij} is the Euclidean distance between city i and city j , $\tau_{ij}(t)$ is pheromone on path $\langle i, j \rangle$ at time t , $\text{tabu}(k)$ is ant k 's tabu list, which is used to record the set of passed through cities, p_{ij} is the transition probability which decides ant to choose the next city j .

From reference [5], [9], [11], the function is defined as follows:

$$p_{ij} = \begin{cases} \frac{[\tau_{ij}]^\alpha \cdot [\eta_{ij}]^\beta}{\sum_{j \in \text{allowed}_k} [\tau_{ij}]^\alpha \cdot [\eta_{ij}]^\beta}, & \text{if } j \in \text{allowed}_k(i) \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

$\text{allowed}_k(i) = \{r - \text{tabu}(k)\}$, which expresses the set of available city to travel; $\eta_{ij} = 1/d_{ij}$ is the distance heuristic function; α is the information heuristic factor; β is the distance heuristic factort.

$$\tau_{ij}(t+n) = (1-\rho) \cdot \tau_{ij}(t) + \Delta \tau_{ij}(t) \tag{12}$$

$$\Delta \tau_{ij}(t) = \sum_{k=1}^m \Delta \tau_{ij}^k(t) \tag{13}$$

ρ is the pheromone volatile coefficient, and $(1-\rho)$ is the residual factor. Usually, it sets $\rho < 1$ to prevent the pheromone infinitely accumulation that will cause the algorithm to fall into local optimum. $\Delta \tau_{ij}^k(0) = 0$.

$$\Delta \tau_{ij}^k = \begin{cases} \frac{Q}{L_k}, & \text{if ant } k \text{ passed path } \langle i, j \rangle \text{ in its tour} \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

This formula is used in ant-cycle models, which is superior to other models [5], [11].

The so-called mobile robot's path planning is to search a collision-free path from the initial state to the goal state according to the request of performance in the certain environment, where includes some obstacles. To realize path planning, it should combine the algorithm and the actual robot [6]. ACO algorithm can apply to the TSP problem directly, however there're some differences between the robot's path planning and the TSP problem [12]. The differences are as follows:

- 1) Ants just are requested to find the shortest path from the starting point to the ending point in robot's path planning, compared with TSP problem ants should travel all cities and then will find the best closed path.
- 2) In robot's path planning, updating the pheromone not only according to the length of the path, but also according to the distance between the path point and the obstacles, compared with TSP problem updating the pheromone according to the length of the path merely.
- 3) In TSP problem, ants are traveling from one city to another city without restrictions of step-size and direction, but in order to realize the path planning more conveniently, every ant should move to one of the 4 adjacent nodes, which are the front, the behind, the left and right, according to the step-size.

3.1 Problem Description and Environment Modeling

As shown in Fig.3, the rectangular global map of the robot's workspace Γ is divided into $m*n$ (length=width=temp) nodes according to the actual situation and robot's movement condition, and temp is the robot's movement step-size namely, which marked black S (x_s, y_s) is the starting point, E (x_e, y_e) is the ending point, between them the light-color regions are the fixed obstacle, the ant needed to find a short and safe path from S to E.

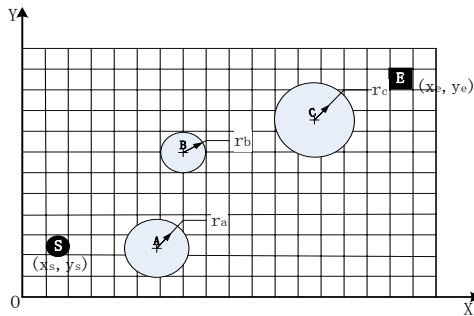


Fig.3. The global map of the robot's workspace

Each node from left to right, from bottom to top is marked: 1, 2, 3,..., i,..., j,..., $m*n$, and each node is indicated by the left bottom coordinate, namely $(x_0, y_0), (x_1, y_1), (x_2, y_2), \dots, (x_i, y_i), \dots, (x_j, y_j), \dots, (x_{m*n-1}, y_{m*n-1})$, then the node i coordinate should satisfy:

$$\begin{cases} x_i = \text{mod} \left[\frac{i-1}{n} \right] \cdot \text{temp} \\ y_i = (\text{int} \left[\frac{i-n-1}{n} \right] + 1) \cdot \text{temp} \end{cases} \tag{15}$$

$\text{mod}[\cdot]$ here represents the residue reduction operation and $\text{int}[\cdot]$ represents the integralized operation. The distance d_{ij} between node i and node j satisfies as follows:

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i + y_j)^2} \tag{16}$$

In order to simplify the computation, the blocking area is three hypothesis circulars, whose radii are r_a , r_b , and r_c , and the centers of circles are A (x_A, y_A), B (x_B, y_B), C (x_C, y_C) respectively. You can judge whether the formula in (17) is established: if establishes, sets the point's coordinate to (∞, ∞) ; otherwise the coordinate is invariable. When generating path, the transition probability formula (11) is according to the path length partially, therefore it may avoid the blocking area effectively.

$$\begin{cases} \sqrt{(x_p - x_a)^2 + (y_p - y_a)^2} - \sqrt{2} r_a \leq 0 \\ \sqrt{(x_p - x_b)^2 + (y_p - y_b)^2} - \sqrt{2} r_b \leq 0 \\ \sqrt{(x_p - x_c)^2 + (y_p - y_c)^2} - \sqrt{2} r_c \leq 0 \end{cases} \tag{17}$$

3.2 Path Point Choice

The ant colony optimization algorithm is a discrete distributed algorithm, the final path is composed with every time sub-path. At time t , ant k must choice the next position depending on the transition probability. Each ant is able to move along up, down, left or right, and without the passed nodes, therefore the feasible region of ant k in position i should satisfies:

$$allowed_k(i) = \left\{ j \mid (j \in \Gamma) \cap (j \notin tabu(k)) \cap (d_{ij} < \sqrt{2}temp) \right\} \tag{18}$$

From the above analysis, the distances d_{ij} between any node i and node j in its feasible region are equal, namely all of η_{ij} are equal, and η_{ij} is insignificant in the in formula (11), therefore η_{ij} in the path planning question should be modified [13] as follows:

$$\eta_{ij}' = \frac{((Maxd_{A(j)E} - d_{jE}) \cdot \omega + \mu)^\gamma}{\sum_{j \in allowed_k} ((Maxd_{A(j)E} - d_{jE}) \cdot \omega + \mu)^\gamma}, j \in allowed_k(i) \tag{19}$$

d_{jE} is the distance from node j to the end point E, $Maxd_{A(j)E}$ is the maximum value of d_{jE} , η_{ij}' can distinguish different node j using this method. It will be more suitable for path planning problem if the distance heuristic function of formula (11) is replaced with formula (19). Parameters(ω, μ, γ) setting must be adjusted conditionally: the temp is more smaller, then the difference among d_{jE} s is very small, these parameters should be increased suitably to distinguish the different nodes; otherwise, its value should be reduced suitably.

3.3 Pheromone Updating

In order to avoid the obstacles safely, it should consider the distance between the path point and the obstacle, and then the security factor λ is introduced, so the pheromone updating formula is modified as follows:

$$\Delta \tau_{ij}^k = \begin{cases} \frac{Q}{F_k}, & \text{if ant } k \text{ passed path } \langle i, j \rangle \text{ in its tour} \\ 0, & \text{otherwise} \end{cases} \quad (20)$$

$F_k = L_k + \lambda \sum_{i=S}^E \frac{3}{d_{iA} + d_{iB} + d_{iC}}$. If λ is bigger, the final path will more far away the obstacle; otherwise, more nearer.

3.4 Algorithm Execution Steps

Step1: Initialization, establishment parameters: NCmax, ant_num, $\tau_{ij}(0) = C$ (C is a constant), $N_c = 0$, $\text{tabu}(k) = \emptyset$, the global best path: $\text{best_path} = \emptyset$, the length of best_path : $\text{min_length} = \infty$

Step2: If $N_c < \text{NCmax}$, transfers to Step3; otherwise transfers to Step7

Step3: Places all the ants on beginning S , namely all ant's tabu list: $\text{tabu}(k) = \{S\}$, and initializes the length of the shortest path in this iteration: $\text{temp_min_length} = \infty$, the best path in this iteration: $\text{temp_path}(N_c) = \emptyset$, the length of path that ant k had passed: $\text{length}(k) = 0$, the number of ants that have found the ending point: $\text{reached_ant} = 0$

Step4: for ($k = 1; k > \text{ant_num}; k++$)

{ant k choices next node j in its feasible region according to formula (11) and (9);

adds j into ant k 's tabu list $\text{tabu}(k)$;

calculates the length of ant k passed path: $\text{length}(k)$;

if ($j == E$)

++ reached_ant ;

if ($\text{reached_ant} == 0$)

transfers to Step6;

else

transfers to Step5;

Step5: for ($i = 0; i < \text{reached_ant}; i++$)

{updating pheromone according to formula (12) and (20);

updating the current length of the shortest path: temp_min_length ;

updating the current best path: $\text{temp_path}(N_c)$;

$\text{min_length} = \text{temp_min_length}$;

$\text{best_path} = \text{temp_path}(N_c)$;

$N_c = N_c + 1$;

for ($k = 1; k > \text{ant_num}; k++$)

{empty tabu list of ant k : $\text{tabu}(k) = \emptyset$;

transfers to Step2;

Step6: for ($k = 1; k > \text{ant_num}; k++$)

{setting the current node of ant k is j ;

transfers to Step4;

Step7: output the global best path and its length, and then algorithm is finished

4 Simulation and Analysis

According to the algorithm execution steps, the simulation results were obtained by using MATLAB. The results are shown in Fig.4. The yellow grid is the starting point, the red one is the ending point, the circular region is the blocking area, and the blue line is the best path generated by this algorithm. The ant number was 20, iterative number $Nc_max=100$, $\alpha=1.4$, $\beta=5$, $\rho=0.5$, $Q=100$. In figure (a), the workspace was divided into $m*n=12*18$ nodes, $temp=5$, $\lambda=5$, $\omega=1$, $\mu=0$, $r=1$; in figure (b) workspace was divided more nodes, $m*n=30*45$, $temp=2$, $\lambda=5$, $\omega=50$, $\mu=10$, $r=10$. Because of the smaller length of step size in figure (b), and the difference of distance between different nodes was reduced, thus the weight of distance heuristic function was increased.

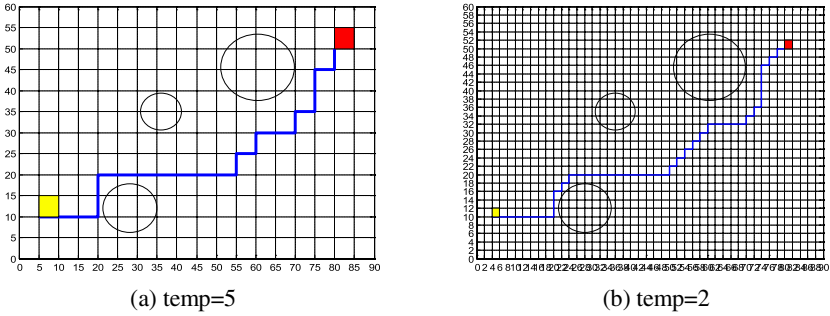


Fig.4. The final best path

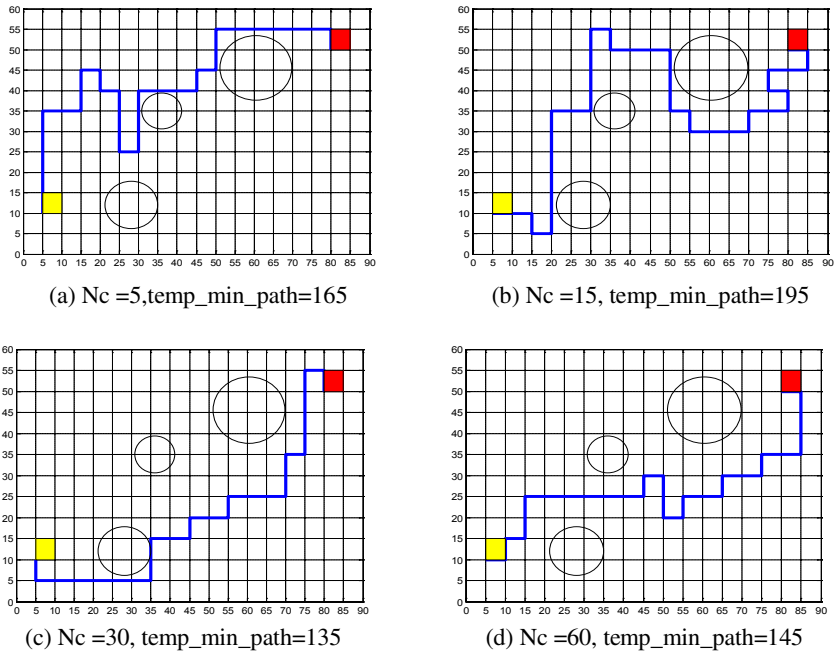


Fig.5. Path generated during the operation, temp=5

Fig.5 is the path generated during the program running. At the beginning moment, because of fewer accumulated pheromone, the generated path was presented bigger randomness and longer length, as shown in figure (a), (b) and (c). Some time after the operation, pheromone was accumulated a lot in some better sub-paths, and the optimal path have restrained to some certain specific sub-paths, but also presented randomness, as shown in figure (d).

5 Conclusions

This article has given the kinematics model for the actual robot, which lays a foundation for further study on motion control and also has modified ACO for mobile robot's path planning, and then operates it through MATLAB. The simulation result has proven the algorithm feasibility and validity. But it has not considered the dynamic obstacles, it will research on the workspace with dynamic in the future, moreover algorithm's convergence and the practical application will also prepare to study. And we will also consider extending this method to other robotic platforms.

References

1. Siegwart, R., Nourbakhsh, I.R.(Write), Li, R.H.(translation): Introduction to Autonomous Mobile Robots. Xi'an jiaotong University Press, Xi'an (2006)
2. Pere: Automatic Planning of Manipulator Movements. *IEEE Trans. on Sys. Man and Cyb.* 11(11), 681–698 (1981)
3. Khatib, O.: Real-time Obstacle Avoidance for Manipulators and Mobile Robots. *Int. J. Robotics Research* 5(1), 90–98 (1986)
4. Holland, J.H.: Genetic Algorithms and the Optimal Allocations of Trails. *SIAM Journal of Computing* 2(2), 88–105 (1973)
5. Colorni, A., Dorigo, M., Maniezzo, V.: Distributed Optimization by Ant Colonies. In: Proceedings of the first European Conference on Artificial Life, Paris, France, pp. 134–142. Elsevier Publishing, Amsterdam (1992)
6. Cui, S.G., Xu, X.L., Lian, Z.G., et al.: Design and Path Planning for a Remote-Brained Service Robot. In: Li, K., Li, X., Irwin, G.W., He, G. (eds.) LSMS 2007. LNCS (LNBI), vol. 4689, pp. 492–500. Springer, Heidelberg (2007)
7. Ding, X.G.: Research on Robot Control. Zhejiang University Press, Hangzhou (2006)
8. Zhang, P.R., Zhang, Z.J., Zheng, X.D., et al.: Design and Realize on Robotic Control System based on 16/32 bit DSP. Tsinghua University Press, Beijing (2006)
9. Dorigo, M., Maniezzo, V., Colorni, A.: The Ant System: Optimizaiton by a Colony of Cooperating Agents. *IEEE Trans. on Sys. Man and Cyb., Part B* 26(1), 1–13 (1996)
10. Dorigo, M., Caro, G.D., Gambardella, L.M.: Ant Algorithms for Discrete Optimization. *Artificial Life* 5(2), 137–172 (1999)
11. Dorigo, M., Stützle, T.(writing), Zhang, J., Hu, X.M., Luo, X.Y., et al.(translation): Ant Colony Optimization. Tsinghua University Press, Beijing (2007)
12. Jin, F.H., Hong, B.R., Gao, Q.J.: Path Planning for Free-flying Space Robot Using Ant Algorithm. *Robot* 24(6), 526–529 (2002)
13. Li, S.Y., et al.: Ant Colony Algorithms with applications. Harbin Institute of Technology Press (2004)

A New Cerebellar Model Articulation Controller for Rehabilitation Robots

Shan Liu¹, Yongji Wang², Yongle Xie¹, Shuyan Jiang¹, and Jinsong Meng¹

¹ School of Automation, University of Electronic Science and Technology of China, Chengdu 610054, China

² Department of Control Science and Engineering, Key Laboratory of Image Processing and Intelligent Control, Huazhong University of Science and Technology, Wuhan 430074, China
shanliu@uestc.edu.cn

Abstract. This paper presents a new cerebellar model articulation controller (CMAC), a sliding-mode-based diagonal recurrent fuzzy CMAC (SDRFCMAC) to robot-assisted rehabilitation for stroke patients. To design the intelligent controller, the CMAC is integrated with some control methods, in which sliding mode technology is used to reduce the dimension of the control system, and fuzzy logic and diagonal recurrent structure is used to solve dynamic problems. The control architecture is represented in terms of stepping optimization system architecture comprising two learning stages to provide robotic assistance for an upper arm rehabilitation task and improve the safety of the human-robot system. Liapunov stability theorem and Barbalat's lemma are adopted to guarantee the asymptotical stability of the system. The effectiveness of the control scheme is demonstrated through a simulated case study.

Keywords: Arm rehabilitation robot, Sliding mode technique, Cerebellar model articulation controller (CMAC), Sliding-mode-based diagonal recurrent fuzzy cerebellar model articulation controller (SDRFCMAC).

1 Introduction

Since the number of patients suffering from stroke is large and the conventional treatment is time consuming, it is a big advance if robots can assist in performing treatment [1]. In the last few years, robot-assisted rehabilitation therapy for the stroke patients has been an active research area, which provides repetitive movement exercise and standardized delivery of therapy with the potential of enhancing quantification of the therapeutic process [2-4].

One of the major difficulties in realizing robot-assisted rehabilitation is the controller design. The existing robotic rehabilitation systems primarily use some conventional controllers to assist the movement of the patient's arms [3-5]. Designing a controller for rehabilitation robots should be difficult, because the external disturbance itself is subjected to another unsolved controller (the human control). In the current work, we consider the affects of the patient in the human-robot rehabilitation system, and then design an interdisciplinary controller with the combination of cerebellar model articulation

controller (CMAC), fuzzy logic, sliding mode technique, and diagonal recurrent network. Note that the presented controller is not specific to a given rehabilitation robot but can be suitable for some kind of exoskeletal wearable arm rehabilitation robots.

This paper is organized as follows. It presents the overall control architecture in Section 2. The conventional controller and the interdisciplinary controller are described in Section 3 and Section 4, respectively. Then the human-robot rehabilitation system is described in Section 5. Simulation results are presented in Section 6. Section 7 discusses the potential contributions and the future work.

2 Control Architecture

The proposed control architecture of the human-robot rehabilitation system is initially presented in the context of testing the potential range of the patient motion, called the test task. In this task, the patients are asked to move their arms to follow the desired joint motion trajectories as accurately as possible. Note that the presented control architecture is not specific to a test task but can be used for any other rehabilitation tasks.

A conventional controller could be used to provide robotic assistance to a patient’s arm movement as and when needed to help him/her to complete the test task. But various robots, different patients, and task-related information may affect the test task. These influential factors may require some adjustments to the accomplishment of the task. As a result, the conventional controller should be aware of these adjustments.

To accommodate the above requirements, an interdisciplinary controller is presented in this work, which learns firstly the basic control ability from the conventional controller, and then provides appropriate assistance for patients based on the available system information and the control effect. Thus, in this work, we take advantage of stepping optimization architecture to train the interdisciplinary controller (Fig. 1). In the course-tuning stage, the interdisciplinary controller is trained to make its output behavior approximate the control surface of the conventional controller. In the fine-tuning stage, the former is further trained to improve the system stability and guarantee the safety of the rehabilitation system.

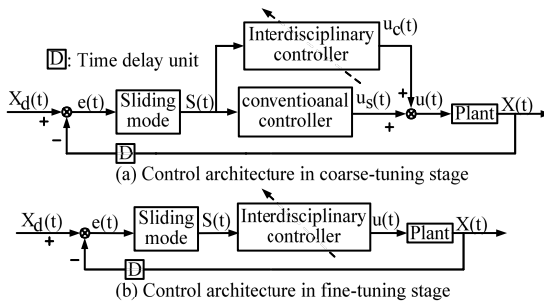


Fig. 1. Control architecture of the human-robot arm rehabilitation system

3 Conventional Controller

Sliding mode control (SMC) is an effective robust control approach for nonlinear multi-input multi-output systems, especially for rigid robots [7][8]. we design a conventional controller based on SMC method to train the interdisciplinary controller.

We define the following sliding mode vector as

$$S = [s_1, s_2, \dots, s_n]^T = CE = [C_1 \ C_2][e_1 \ e_2]^T . \quad (1)$$

where $C_1 = \text{diag}(c_1, \dots, c_n) \in R^{n \times n}$ is a positive matrix, $C_2 \in R^{n \times n}$ is a unit matrix, $E = X_d - X$ is the tracking error vector. The SMC law is designed as

$$u = \text{inv}(B(X))(-F(X) + \dot{x}_{2d} + C_1(\dot{x}_{1d} - \dot{x}_1) + D\text{sat}(S) + \dot{S} + C_1\text{sat}(S)) . \quad (2)$$

where $\text{inv}(B(X))$ represents the inverse of the matrix $B(X)$, $\text{sat}(s)$ is a saturation function, which is defined as follows:

$$\text{sat}(s) = \begin{cases} 1 & s/\delta > 1 \\ s/\delta & -1 \leq s/\delta \leq 1 \\ -1 & s/\delta < -1 \end{cases} . \quad (3)$$

where $\delta > 0$ is the layer thickness, $|s/\delta| < \kappa$ is assumed.

The stability of the conventional controller is guaranteed through choosing a Liapunov function $V_1 = \frac{1}{2}(S - \text{sat}(S)\delta)^T(S - \text{sat}(S)\delta)$ based on Liapunov stability theory.

4 Interdisciplinary Controller

In this section, we first present the model of the interdisciplinary controller, followed by the training details and stability analysis of the interdisciplinary controller.

4.1 Model

CMAC, a non-fully connected associative memory network, has good generalization capability and fast learning property. Some applications of CMAC for complex dynamic systems have been presented in [9][10][11][12]. The interdisciplinary controller is designed as a modified CMAC with the combination of fuzzy logic, sliding mode technique, and diagonal recurrent network, called as sliding mode based diagonal recurrent fuzzy CMAC (SDRFCMAC). The model of the SDRFCMAC is shown in Fig. 2, in which T denotes the delay time. The network is composed of input space, association memory space with recurrent units, receptive field space, weight memory space and output space. The input of the SDRFCMAC is the signed sliding mode vector S through (1). Firstly, $S \in R^n$ is normalized, the input space is quantized into discrete regions (called elements), and the number of elements is Ne . Next, S is mapped into the association memory space through receptive basis functions, where

the space consists of $n \times Na$ blocks, a complete block is formed by Nr elements, Na is the number of blocks relative to each input. Thirdly, the association memory matrix $A \in R^{n \times Na}$ is mapped into the receptive field space through multidimensional receptive field functions. Lastly, the receptive field vector $Rs \in R^{Nr}$ is projected onto weight matrix $W \in R^{m \times Nr}$ to computer the output $Y \in R^m$. The SDRFCMAC consists of two primary functions that are performed in the association memory space and the receptive field space, respectively.

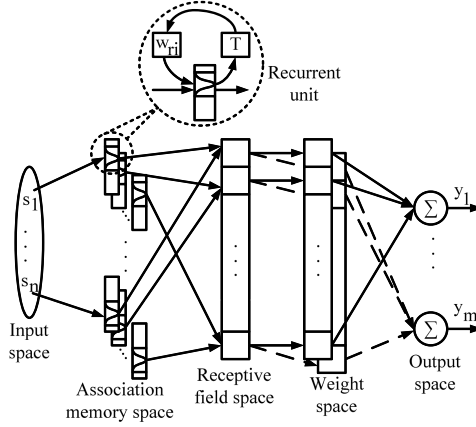


Fig. 2. Illustration of the SDRFCMAC model

1) Receptive basis function: Gaussian function is adopted here, which can be represented as

$$\alpha_{ij}^h = \exp\left(-\frac{(s_{rij} - \lambda_{ij})^2}{\sigma_{ij}^2}\right), i = 1, \dots, n; h = 1, \dots, Nr; j = 3p + h \quad (4)$$

where α_{ij}^h represents the j^{th} block of the i^{th} input s_i with the mean λ_{ij} and variance σ_{ij} in the h^{th} layer, and $p = 0, \dots, \text{ceil}\left[\frac{Ne-h}{Nr}\right]$. The mean and variance can be expressed in the vector form $\Lambda \in R^{n \times Na}$ and $\Sigma \in R^{n \times Na}$. The input s_{rij} in time step k is represented as

$$s_{rij}(k) = s_i(k) + w_{rij} \alpha_{ij}^h(k-T) \quad (5)$$

where w_{rij} is the recurrent weight, $\alpha_{ij}^h(k-T) \triangleq \alpha_{ijT}^h$ denotes the value of $\alpha_{ij}^h(k)$ through time delay T . The recurrent weight matrix can be expressed as $Wr \in R^{n \times Na}$.

2) Multidimensional receptive field function: Receptive fields are formed by blocks. The multidimensional receptive field function is defined as

$$r_h = \prod_{i=1}^n \sum_{j=1}^{3p+h} \alpha_{ij}^h, \quad p = 0, 1, \dots, \text{ceil}\left[\frac{Ne-h}{Nr}\right], \quad h = 1, 2, \dots, Nr \quad (6)$$

In this SDRFCMAC scheme, no receptive field is formed by the combination of blocks in different layers. Thus, the number of receptive fields is Nr . This kind of composition reduces the memory requirement, and makes nearby inputs can produce similar outputs, which provide local generation to SDRFCMAC.

The output of the SDRFCMAC is expressed as $Y = W \cdot R_s$, where the l^{th} element in Y is $y_l = \sum_{h=1}^{Nr} w_{lh} r_h$.

4.2 Learning Process

The learning process of the SDRFCMAC includes two stages, the coarse-tuning stage and the fine-tuning stage.

4.2.1 Coarse-Tuning Stage

The purpose of this stage is to enable the output behaviour of the SDRFCMAC to approximate the control surface of the SMC controller. The control system is shown in Fig. 1(a). The control law $u(t)$ is used as the target output; the error function is defined as

$$E_1(k) = \frac{1}{2} (u_c(k) - u(k))^T (u_c(k) - u(k)) = \frac{1}{2} \sum_{l=1}^m u_{sl}(k)^2. \quad (7)$$

Initially, the SDRFCMAC weight matrixes are set as zero matrixes. And then, according to the gradient descent method, these weight matrixes are updated at each time step by the learning rules (8).

$$\begin{cases} \dot{w}_{lh} = -\eta_{w1} \partial E_1 / \partial w_{lh} = \eta_{w1} u_{sl} r_h \\ \dot{\lambda}_{ij} = -\eta_{\lambda1} \partial E_1 / \partial \lambda_{ij} = \eta_{\lambda1} \sum_{l=1}^m (u_{sl} w_{lh}) \partial r_h / \partial \lambda_{ij} \\ \dot{\sigma}_{ij} = -\eta_{\sigma1} \partial E_1 / \partial \sigma_{ij} = \eta_{\sigma1} \sum_{l=1}^m (u_{sl} w_{lh}) \partial r_h / \partial \sigma_{ij} \\ \dot{w}_{rij} = -\eta_{r1} \partial E_1 / \partial w_{rij} = -\eta_{r1} \sum_{l=1}^m (u_{sl} w_{lh}) \partial r_h / \partial w_{rij} \end{cases}. \quad (8)$$

where η_{w1} , $\eta_{\lambda1}$, $\eta_{\sigma1}$, and η_{r1} are positive constants, the subscript h can be derived from the subscript j , and

$$\begin{cases} \frac{\partial r_h}{\partial \lambda_{ij}} = 2 \left(\prod_{k=1, \neq i}^n \sum_{j=1}^{3p+h} \alpha_{kj}^h \right) \alpha_{ij}^h \frac{(s_{rij} - \lambda_{ij})}{\sigma_{ij}^2} \\ \frac{\partial r_h}{\partial \sigma_{ij}} = 2 \left(\prod_{k=1, \neq i}^n \sum_{j=1}^{3p+h} \alpha_{kj}^h \right) \alpha_{ij}^h \frac{(s_{rij} - \lambda_{ij})^2}{\sigma_{ij}^3} \\ \frac{\partial r_h}{\partial w_{rij}} = -2 \left(\prod_{k=1, \neq i}^n \sum_{j=1}^{3p+h} \alpha_{kj}^h \right) \alpha_{ij}^h \alpha_{ijT}^h \frac{(s_{rij} - \lambda_{ij})}{\sigma_{ij}^2} \end{cases}. \quad (9)$$

The gradient descent method can guarantee the convergence of the parameters λ_{ij} , σ_{ij} , and w_{rij} , and the output of the receptive field basis functions are limited in $[0,1]$. Therefore, the stability of the control system will not be destroyed due to the adaptive learning rules shown in (8).

4.2.2 Fine-Tuning Stage

The objective of this stage is to improve the system stability. The control system is shown in Fig. 1(b). Learning rules are derived from the gradient of $SS^{\dot{}}$ with respect to the parameters in the SDRFCMAC.

$$\begin{cases} \dot{w}_{lh} = -\eta_{w2} \partial SS^{\dot{}} / \partial w_{lh} = \eta_{w2} \sum_{i=1}^n (s_i b_{il}) r_h \\ \dot{\lambda}_{ij} = -\eta_{\lambda2} \partial SS^{\dot{}} / \partial \lambda_{ij} = \eta_{\lambda2} \sum_{i=1}^n \sum_{l=1}^m (s_i b_{il} w_{lh}) \partial r_h / \partial \lambda_{ij} \\ \dot{\sigma}_{ij} = -\eta_{\sigma2} \partial SS^{\dot{}} / \partial \sigma_{ij} = \eta_{\sigma2} \sum_{i=1}^n \sum_{l=1}^m (s_i b_{il} w_{lh}) \partial r_h / \partial \sigma_{ij} \\ \dot{w}_{rij} = -\eta_{r2} \partial SS^{\dot{}} / \partial w_{rij} = \eta_{r2} \sum_{i=1}^n \sum_{l=1}^m (s_i b_{il} w_{lh}) \partial r_h / \partial w_{rij} \end{cases} \quad (10)$$

where η_{w2} , $\eta_{\lambda2}$, $\eta_{\sigma2}$ and η_{r2} are positive constants. The parameters update equations are given by

$$\begin{cases} W(k+1) = W_{coarse-tuning}(k) + \dot{W} \\ \Lambda(k+1) = \Lambda_{coarse-tuning}(k) + \dot{\Lambda} \\ \Sigma(k+1) = \Sigma_{coarse-tuning}(k) + \dot{\Sigma} \\ Wr(k+1) = Wr_{coarse-tuning}(k) + \dot{Wr} \end{cases} \quad (11)$$

where $W_{coarse-tuning}$, $\Lambda_{coarse-tuning}$, $\Sigma_{coarse-tuning}$, and $Wr_{coarse-tuning}$ are the final SDRCMAC parameters at the coarse-tuning stage; the behavior of the sliding mode controller is implicit in these parameters. Since the learning error will not accumulated in the fine-tuning stage, the instability caused by the continued learning after the tracking error has been reduced can be solved by (11).

4.3 Stability Analysis

In the coarse-tuning stage, the SMC law (2) can guarantee the stability of the control system. In the following, the stability in the fine-tuning stage will be proved.

For the stability analysis, we assume the optimal parameter matrixes \bar{W} , $\bar{\Lambda}$, $\bar{\Sigma}$, and \bar{Wr} exists, which makes the SDRFCMAC output to approximate the SMC law (2) with an error smaller than ξ , ξ is a positive number.

$$\max(\bar{u}(S, \bar{W}, \bar{\Lambda}, \bar{\Sigma}, \bar{Wr}) - u_s) < \xi \quad (12)$$

where $\bar{u}(S, \bar{W}, \bar{\Lambda}, \bar{\Sigma}, \bar{W}r) \triangleq \bar{W}\bar{R}s$, Then, $u_s = \bar{W}\bar{R}s + \Xi$, where $\Xi \in R^m$. According to (2) and (12), the following equation can be derived

$$\begin{aligned} \dot{S} &= -\frac{1}{2}(Dsat(S) + C_1sat(S) + d(t)) + B(u_s - u) \\ &= -\frac{1}{2}(Dsat(S) + C_1sat(S) + d(t)) + B(\tilde{W}\hat{R}s + \hat{W}\tilde{R}s + \tilde{W}\tilde{R}s) \end{aligned} \quad (13)$$

where $\tilde{W} = \bar{W} - \hat{W}$ and $\tilde{R}s = \bar{R}s - \hat{R}s$. Taylor linearization technique is employed to transform the nonlinear function into a partially linear form

$$\begin{aligned} \tilde{R}s &= \left[\frac{\partial R_s}{\partial \Lambda} \right]_{\Lambda=\hat{\Lambda}} \tilde{\Lambda} + \left[\frac{\partial R_s}{\partial \Sigma} \right]_{\Sigma=\hat{\Sigma}} \tilde{\Sigma} + \left[\frac{\partial R_s}{\partial W_r} \right]_{W_r=\hat{W}r} \tilde{W}r + H \\ &= \left[tr(\tilde{\Lambda}^T r_{h\Lambda}|_{\Lambda=\hat{\Lambda}}) \right] + \left[tr(\tilde{\Sigma}^T r_{h\Sigma}|_{\Sigma=\hat{\Sigma}}) \right] + \left[tr(\tilde{W}r^T r_{hW_r}|_{W_r=\hat{W}r}) \right] \\ &= R_\Lambda + R_\Sigma + R_{W_r} + H \end{aligned} \quad (14)$$

where $\tilde{\Lambda} = \bar{\Lambda} - \hat{\Lambda}$, $\tilde{\Sigma} = \bar{\Sigma} - \hat{\Sigma}$, $\tilde{W}r = \bar{W}r - \hat{W}r$, H is a high-order term, and $R_\Lambda, R_\Sigma, R_{W_r} \in R^{N_r}$.

Choose the Liapunov function as

$$V_2 = S^T S + tr(\tilde{W}^T \tilde{W})/\eta_{w2} + tr(\tilde{\Lambda}^T \tilde{\Lambda})/\eta_{\lambda2} + tr(\tilde{\Sigma}^T \tilde{\Sigma})/\eta_{\sigma2} + tr(\tilde{W}r^T \tilde{W}r)/\eta_{r2} \quad (15)$$

Differentiating (15) with respect to time and using (11), (13) and (14) yields

$$\dot{V}_2 = -S^T(C_1sat(S) + Dsat(S) + d(t)) + S^T B\Delta \quad (16)$$

where $\Delta = \hat{W}H + \tilde{W}\tilde{R}s + \Xi$ is assumed to be bounded by $\|B\Delta\| \leq \kappa\|C_1\| < \|C_1\|$. Then,

$$\dot{V}_2 \leq -S^T C_1sat(S) + \|S\|\|B\Delta\| - S^T Dsat(S) + \|S\|\|d(t)\| \quad (17)$$

In the case of $|S/\delta| > 1$, $\dot{V}_2 \leq -\|S\|(\|C_1\| - \|B\Delta\|) - \|S\|(D - \|d(t)\|) < 0$; in the case of $|S/\delta| < 1$, $\dot{V}_2 \leq -\|S\|(\kappa\|C_1\| - \|B\Delta\|) - \|S\|(\kappa D - \|d(t)\|) \leq 0$. Since \dot{V}_2 is negative semidefinite, that is $V_2 \leq V_2(0)$, it implies that S , \tilde{W} , $\tilde{\Lambda}$, $\tilde{\Sigma}$, and $\tilde{W}r$ are bounded. Let function

$$L \equiv S(\kappa\|C_1\| - \|B\Delta\|) + S(\kappa D - \|d(t)\|) \leq \|S\|(\kappa\|C_1\| - \|B\Delta\| + \kappa D - \|d(t)\|) \leq -\dot{V}_2 \quad (18)$$

and integrate L with respect to time, it can be shown that $\int_0^t L d\tau \leq V_2 - V_2(0)$. Because $V_2(0)$ is bounded and V_2 is non-increasing and bounded, the following result can be shown $\lim_{t \rightarrow \infty} \int_0^t L d\tau < \infty$. In addition, since \dot{L} is bounded by Barbalat's lemma, it

can be shown that $\lim_{t \rightarrow \infty} L = 0$. That is, $S \rightarrow 0$ as $t \rightarrow \infty$. As a result, the control system is asymptotically stable.

5 Simulation Results

A device for Robotic Assisted Upper Extremity Repetitive Therapy (RUPERT™ IV) is used as the main hardware platform in this work, which uses four pneumatic muscles to actuate shoulder elevation, elbow extension, supination, and wrist extension [3]. Depending on the motion ability test of these patients' arms, the functions of the active joint torques are designed.

The dynamic model of the human-robot rehabilitation system is designed based on Lagrange equation in Robotics (given in (19)) [6].

$$M(q)\ddot{q} + C(q, \dot{q})\dot{q} + Kq + B\dot{q} + G(q) + \Delta(q, \dot{q}) = \tau \quad (19)$$

where, $q, \dot{q}, \ddot{q} \in R^4$ is the joint angle, joint angle velocity, and angle acceleration vector, respectively; $M(q) \in R^{4 \times 4}$ is the inertia matrix; $C(q, \dot{q}) \in R^4$ is the vector of Coriolis and centrifugal forces; $K \in R^{4 \times 4}$ is the joint friction matrix, $B \in R^{4 \times 4}$ is the joint viscosity matrix; $G(q) \in R^4$ is the gravity vector; $\Delta(q, \dot{q})$ is the error of the model; τ includes the patient's active torque and the control signal.

In the simulation, the control objective is to let the system state X track the reference trajectory X_d . The SDRFCMAC used in these systems is characterized as follows: $N_e = 4$; $N_r = 3$; $N_a = 6$. The receptive fields are selected to cover the input space $\{[-1, 1], [-1, 1]\}$ along each input dimension. Therefore, the initial values of the parameters for the receptive field basis functions in the coarse-tuning stage are $[\lambda_{r1}, \lambda_{r2}, \lambda_{r3}, \lambda_{r4}, \lambda_{r5}, \lambda_{r6}] = [-1.25, -0.75, -0.25, 0.25, 0.75, 1.25]$, and $\sigma_{ij} = 0.75$. The initial weight W and W_r in the coarse-tuning stage are set as zero matrixes, and the final parameters in the coarse-tuning stage are chosen as the initial parameters in the fine-tuning stage.

The initial state are $q_0 = [0, 0, 0, 0]^T$, $\dot{q}_0 = [0, 0, 0, 0]^T / s$, and the reference trajectories are set as $q_d(t) = [30 + 10 \sin t, 60 + 30 \sin t, 45 \sin t, 15 + 45 \sin t]^T$. The control parameters are chosen as $\delta = 0.3$, $C_1 = \text{diag}(10, 10)$, $\kappa = 0.3$, $D = 5$, $\eta_{wi} = 0.6$, $\eta_{\lambda i} = \eta_{\sigma i} = \eta_{r i} = 0.1$, where $i = 1, 2$. These parameters are chosen through trial and error to achieve satisfactory performance. For comparison, the fuzzy CMAC (FCMAC) [10], the SMC [7] and the SDRFCMAC are used in the simulation. The shoulder and elbow flexure/extension joint motion is more influential in the upper arm movements. Thus, Fig. 3 shows the angle tracking responses of these three methods in these two joints. The results of the SDRFCMAC and the SMC scheme are all satisfactory, and slightly better than that of the FCMAC. For further comparing SMC and SDRFCMAC, Fig. 4 shows the angle velocity tracking responses these two methods. These figures illustrate that the SDRFCMAC controller can more smoothly track the reference trajectory.

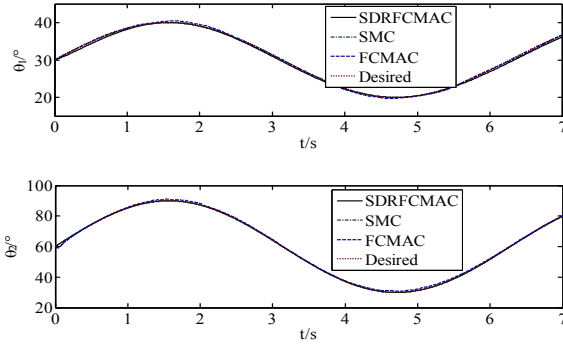


Fig. 3. Angle tracking responses of the human-robot system

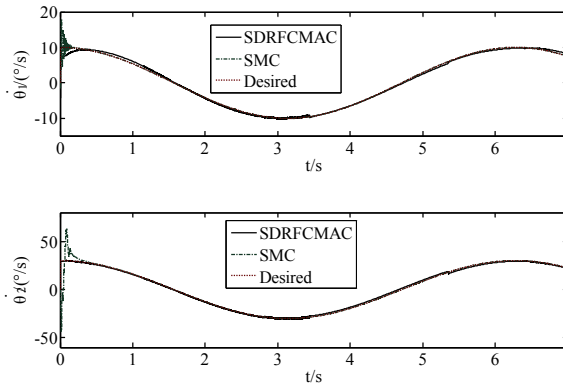


Fig. 4. Angle velocity tracking responses of the human-robot system

6 Conclusions and Future Work

In this paper, we present a new control approach for the human-robot rehabilitation system that includes the coordination between an interdisciplinary controller and a conventional controller, which are used to provide robotic assistance for an upper arm rehabilitation task. According to Liapunov stability theorem and Barbalat's lemma, the asymptotical stability of the human-robot system is guaranteed. The simulation results demonstrated the usefulness of the controller in the test rehabilitation task.

An important direction for future development involves testing the usability of the proposed control method with stroke patients in laboratorial and clinical experiments. New methods to detect human state information can be integrated into the control system. For instance, Electromyogram (EMG) signals can be used to monitor the patient's muscle state to detect their exhaustion.

Acknowledgments. This work is supported by grants from the National Nature Science Foundation of China, No 60674105. The authors would like to thank Prof. Jiping He for the guidance regarding the RUPERT modeling and control.

References

1. Lum, P.S., Burgar, C.G., Shor, P.C., Majmundar, M., Van der Loos, H.F.M.: Robot-assisted Movement Training Compared with Conventional Therapy Techniques for the Rehabilitation of Upper-limb Motor Function after Stroke. *Arch. Phys. Med. Rehab.* 83, 952–959 (2002)
2. Kahn, L.E., Lum, P.S., Rymer, W.Z., Reinkensmeyer, D.J.: Robot-assisted Movement Training for the Stroke-impaired Arm: Does it Matter What the Robot Does? *J. Rehab. Res. Dev.* 43, 619–630 (2006)
3. Sugar, T.G., He, J., Koeneman, E.J., Koeneman, J.B., Herman, R., Huang, H., Schultz, R.S., Herring, D.E., Wanberg, J., Balasubramanian, S., Swenson, P., Ward, J.A.: Design and Control of RUPERT: a Device for Robotic Upper Extremity Repetitive Therapy. *IEEE Trans. Neural Sys. & Rehab. Eng.* 15, 336–346 (2007)
4. Krebs, H.I., Palazzolo, J.J., Dipietro, L., Ferraro, M., Krol, J., Rannekleiv, K., Volpe, B.T., Hogan, N.: Rehabilitation Robotics: Performance-based Progressive Robot-assisted Therapy. *Auto. Robots.* 15, 7–20 (2003)
5. Ju, M.S., Lin, C.C.K., Lin, D.H., Hwang, I.S., Chen, S.M.: A Rehabilitation Robot With Force-Position Hybrid Fuzzy Controller: Hybrid Fuzzy Control of Rehabilitation Robot. *IEEE Trans. Neural Sys. & Rehab. Eng.* 13, 349–358 (2005)
6. Craig, J.J.: *Introduction to Robotics: Mechanics and Control*. Prentice Hall, New Jersey (2004)
7. Liu, S., Wang, Y., Fang, H., Xu, Q.: Trajectory Tracking Sliding Mode Control for Robot. *Microcomputer Inf.* 24, 261–262 (2008)
8. Liu, J.K.: *MATLAB Simulation for Sliding Mode Control*. Tsinghua Press, Beijing (2005)
9. Chiang, C.T., Lin, C.S.: CMAC with general basis functions. *Neural Network* 9, 1199–1211 (1996)
10. Sun, W., Wang, Y.N.: Fuzzy Cerebellar Model Articulation Controller and its Application on Robotic Tracking Control. *Control Theory & Application* 23, 38–42 (2006)
11. Lin, C.M., Chen, L.Y., Chen, C.K.: RCMAC Hybrid Control for MIMO Uncertain Nonlinear Systems Using Sliding-mode Technology. *IEEE Trans. Neural Networks* 18, 708–720 (2007)
12. Yeh, M.F.: Single-input CMAC Control System. *Neurocomputing* 70, 2638–2644 (2007)

Layer-TERRAIN: An Improved Algorithm of TERRAIN Based on Sequencing the Reference Nodes in UWSNs

Yue Liang and Zhong Liu

Naval University of Engineering, Wuhan 430033, China
liangyue0220@gmail.com

Abstract. The Layer-Terrain algorithm is proposed in the background of underwater environment, focusing on reduction of the accumulated location error. Based on the TERRAIN algorithm, this algorithm introduces the concept of layer, sequences the reference nodes according to the information of layer and estimates the position by maximum likelihood estimation. This sequencing process not only makes use of redundant information, but also limits the amount of computing data. The simulation experiments show that the improved algorithm efficiently deals with error propagation among the network and improves the localization precision of TERRAIN algorithm.

Keywords: Node localization, Underwater wireless sensor networks, TOA, Layer, Layer-TERRAIN.

1 Introduction

1.1 Location in UWSNs

Wireless sensor networks are composed of large numbers of sensor nodes that are scattered in the region of interest to acquire some physical data. These sensor nodes should have the ability of sensing, processing and communicating ^[1].

Many applications of wireless sensor networks require the location information of sensor nodes, such as environment monitoring, remote controlling, target tracking, coverage and routing.

Lots of localization algorithms for wireless sensor networks have been proposed. These algorithms are divided into two categories: range-based algorithm ^[2-5] and range-free algorithm ^[6-10]. The former is defined by protocols that use absolute point-to-point distance or angle estimates for calculating location. The methods of distance or angle estimates include time of arrival (TOA), time difference of arrival (TDOA), received signal strength indicator (RSSI) and angle of arrival (AOA). The latter makes use of the information of connectivity and estimated distance for calculating location. This paper concentrates on the range-based.

The underwater wireless sensor network is a newly application domain ^[11]. Because of complicated physical environment and limited communication bandwidth in the ocean, the sound signal is chosen as the information transmission medium. The method

of TOA is used to estimate the absolute point-to-point distance. The node localization process of UWSN is depicted as follow: first, upgrading the normal nodes to anchor nodes by some automatic underwater vehicle (AUV) or other moving devices, second, localizing every node of the UWSN by iterative algorithm. However, the iterative algorithm leads to error accumulation problem when measurement error exists. When networks are large, while the anchor nodes are not enough, the results of localization are often unacceptable.

There are two ways that deal with the problem of error accumulation: repeatedly measurements^[12] and circulating refinements^[13]. The former needs many times measurements of distances, the latter yields large amount cost of communication and computation, sometimes, the algorithm is uncertain because of un-prediction of cycle index^[14].

Therefore, we need a more efficient method to deal with the error accumulation problem of the iterative algorithm.

1.2 TERRAIN

TERRAIN is a range-based algorithm proposed by Savarese. This algorithm consists of three steps:

1. Set up relative coordinate system based on each anchor node.
2. Measure distance between anchor nodes and unknown nodes.
3. Compute the geographic position by maximum likelihood estimates.

In step 1, the relative coordinate system is set up with the anchor nodes as the origin. The assumption based coordinates (ABC) algorithm is used to localize the initial reference nodes. Three initial reference nodes have to be localized in two-dimension space and four in three-dimension space. In this paper, we discuss the two-dimension space.

In step 2, based on the initial references nodes, the maximum likelihood estimate is used to localize the relative coordinate of unknown nodes with the iterative algorithm. After that, the distances between nodes and origin are obtained.

In step 3, when one node computes more than three distances, the maximum likelihood estimate is used to compute its geographic position.

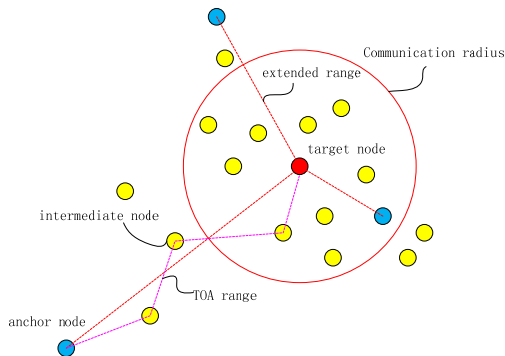


Fig. 1. The TERRAIN algorithm

1.3 Our Work

In the second step, TERRAIN estimates the node position without sequencing the reference nodes. When the reference nodes have large error, the estimation process related to these nodes definitely yields larger error and leads to error accumulation.

Therefore, we improve the TERRAIN algorithm in two ways:

1. Introduce the concept of layer, and sequence the reference nodes mainly according to the information of layer.
2. Sequence the reference nodes according to the acoustic propagation time if they are in the same layer.

The rest of this paper is organized as follow. In section 2, we propose the layer-TERRAIN algorithm. In section 3, the experiment results are given and analyzed. Section 4 concludes the paper and outlines the future works.

2 Layer-TERRAIN Algorithm

2.1 Definition of Layer

Layer indicates the node position precision indirectly. In the relative coordinate system, the transmission form of information between nodes is described in table 1.

Table 1. Transmission form of information

ID	Layer	X	Y	Sen_time	Rec_time
----	-------	---	---	----------	----------

The symbols that used in transmission form are explained as follow:

- ID- the unique identifier of each node.
- Layer- the layer of each node.
- X-the relative X-axis coordinate.
- Y-the relative Y-axis coordinate.
- Sen_time - the time that a node sends the information. In the premise of time synchronization, $T_{\text{communicate}}$ refers the maximum communication time, and T_{compute} refers the maximum computation time. Time between two iterative process is defined as $T = T_{\text{communicate}} + T_{\text{compute}}$, and $\text{Sen_time} = n * \Delta T$, in which n is a positive integer. After sending the information, the node switches to the sleep mode.
- Rec_time - the time that a node receives the information.
- The definition of layer is on the basis of the following rules:

- The layer of the initial reference nodes is zero. In other words, the layer of reference nodes determined by the ABC algorithm is zero.
- The layer of the node that uses the initial reference nodes to estimate its position is one. The layer of the node that uses the reference nodes in the first layer to estimate its position is two, and the like.
- The layer of the node that uses the reference nodes in different layers to estimate its position is adding 1 to the least layer of the reference nodes.

2.2 Sequencing of Reference Nodes

Assuming that a node receives k reference nodes information, the process of sequencing the reference nodes is shown as follow:

1. If $k < 3$, turn to step 2. If $k \geq 3$, turn to step 3.
2. Receive new information of reference nodes within another ΔT .
3. If $k \leq 10$, turn to step 5. If $k > 10$, turn to step 4.
4. Sequence the newly added reference nodes within ΔT .
5. The maximum likelihood estimate is used to estimate node position.

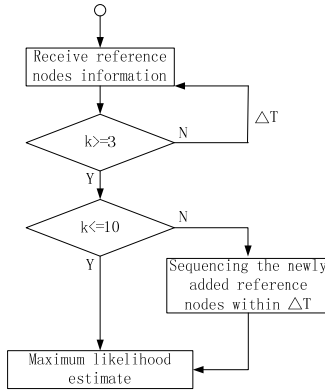


Fig. 2. Sequencing the reference nodes

The pseudo codes are given as follow.

```

while (the node is not located)
{
    if k < 3
        waiting for new reference nodes within ΔT
    continue
}
    
```

```

else if k>10
    sequencing the newly added reference nodes,
    and throw the unwanted ones
else
    the maximum likelihood estimate is used to
    estimate node position
}

```

2.3 Maximum Likelihood Estimation

Reference demonstrates that the precision of node position is not improved apparently when the number of reference nodes is more than ten, however, extra reference nodes lead to extra cost of computing. Therefore, in this paper, we enact that the maximum number of reference nodes involved in computing is ten. This rule not only makes use of redundant information, but also limits the amount of computing information.

Let (x, y) refers to the coordinate of the unknown node, $(x_1, y_1) \cdots (x_{10}, y_{10})$ refers to coordinates of the sequenced reference nodes and $(r_1, r_2 \cdots r_{10})$ refers to the distances between the reference nodes and unknown node. The linear equation is $AX = B$

$$A = \begin{bmatrix} 2(x_1 - x_{10}) & 2(y_1 - y_{10}) \\ \vdots & \vdots \\ 2(x_9 - x_{10}) & 2(y_9 - y_{10}) \end{bmatrix} \quad (1)$$

$$B = \begin{bmatrix} x_1^2 - x_{10}^2 + y_1^2 - y_{10}^2 + r_1^2 - r_{10}^2 \\ \vdots \\ x_9^2 - x_{10}^2 + y_9^2 - y_{10}^2 + r_9^2 - r_{10}^2 \end{bmatrix} \quad (2)$$

The position of unknown node estimated by maximum likelihood estimate is $\hat{X} = (A^T A)^{-1} A^T B$.

2.4 Localization Process of UWSN

After above analysis, the node localization in UWSN is depicted as follow:

- AUV or other moving devices localize a small fraction of nodes as anchor. Relative coordinate systems are set up with the anchor nodes as origin.
- In the relative coordinate systems, an optimized iterative algorithm based on the information of layer is used to estimate the node position. After that, the distance between the nodes and origins are obtained.
- The maximum likelihood estimate is used to localization the node geographic position.

3 Experiment and Analysis

3.1 Analysis of Algorithm

In this set of experiment, many nodes from 100 to 500 are placed randomly in a 20*20 square. Each node has the same communication radius of 4. The proportion of the anchor nodes is 5% and the measurement error is 5% of radius. Figure 3 shows the relation between position error and number of nodes. Figure 4 shows the relation between position error and number of anchor nodes. Figure 5 shows the relation between the position error and connectivity. The position error of TERRAIN algorithm is about 39% under the usual conditions. From these figures, we can conclude that the layer-TERRAIN algorithm is more precious than TERRAIN algorithm.

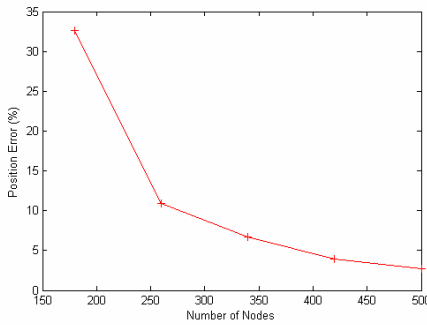


Fig. 3. Relation between position error and number of nodes

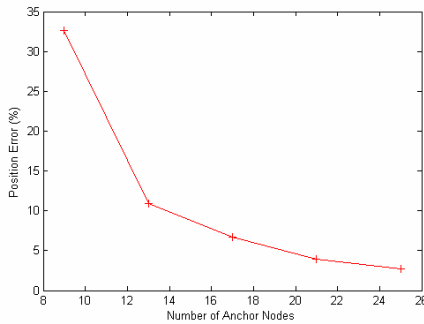


Fig. 4. Relation between position error and number of anchor nodes

3.2 Analysis of Accumulation Error Controlling

In order to confirm the effects on error accumulation controlling, we do another experiment. 400 nodes are randomly placed in a 40*40 square. Each node has the same

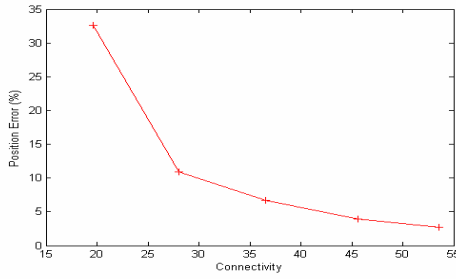


Fig. 5. Relation between position error and connectivity

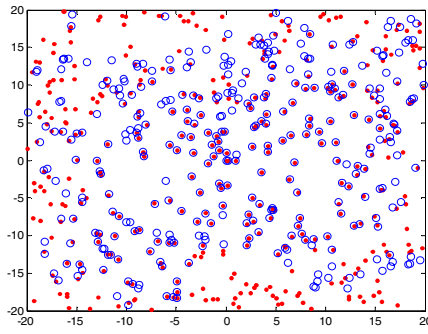


Fig. 6. Real position and estimated position after error accumulation controlling

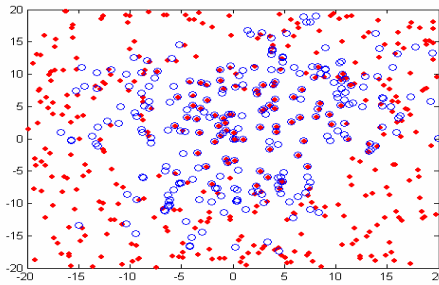


Fig. 7. Real position and estimated position without error accumulation controlling

radius of 4 and the measurement error is 5% of radius. The anchor nodes are placed in (0, 0), (0, 1) and (1, 1). The blue ‘o’ stands for the real node position. The red ‘.’ stands for the estimated node position. Figure 6 shows the node real position and estimated position after error accumulation controlling. Figure 7 shows the node real position and estimated position without error accumulation controlling. The experiment demonstrates that after

error accumulation controlling, 65.01% of nodes are estimated, while without error accumulation controlling only 33.00% of nodes can be estimated. Layer-TERRAIN improves the number of nodes that can be estimated.

4 Conclusion

We present an improved localization algorithm for underwater wireless sensor networks. Through our study, we find that the best method of ranging in ocean environment is TOA or TDOA if range-based localization method is used. Furthermore, we present the layer-TERRAIN algorithm to deal with the measurement errors and error accumulation. The experiment demonstrates that the improved algorithm has better performance than the TERRAIN.

Acknowledgments. The author would like to thank the reviewers whose suggestions have improved the quality of the paper. This research is funded by National Defence Research Foundation of China under Grant Number 513040203.

References

1. Sun, L.M., Li, J.Z., Chen, Y., Song, H.S.: *The Wireless Sensor Network*. Tsinghua University Press, Beijing (2005)
2. Girod, L., Estin, D.: Robust Range Estimation Using Acoustic and Multimodal Sensing. In: 2001 IEEE/RSJ International Conference on Intelligent Robots and System, Hawaii, USA, pp. 1312–1320 (2001)
3. Andreas, S., Han, C.C., Mani, B.: Dynamic Fine-Grained Location in Ad-hoc Networks of Sensors. In: 7th Annual International Conference on Mobile Computing and Network, Rome, Italy, pp. 166–179 (2001)
4. Priyantha, N.B., Chakroborty, A., Balakrishnan, H.: The Cricket Location-Support System. In: 6th Annual International Conference on Mobile Computing and Network, Boston, USA, pp. 32–43 (2001)
5. Savarese, C., Rabay, J.M., Beutel, J.: Locating in Distributed Ad-Hoc Wireless Sensor Network. In: 2001 IEEE International Conference on Acoustics, Speech and Signal, Salt Lake, pp. 2037–2040 (2001)
6. Nicolescu, D., Nath, B.: Ad-Hoc Positioning System. In: 2001 Global Telecommunications Conference, San Antonio, pp. 2926–2931 (2001)
7. Nicolescu, D., Nath, B.: DV-Based Positioning in Ad-Hoc Networks. *Telecommunication System* 22, 267–280 (2003)
8. Nagpal, R.: Organizing a Global Coordinate System from Local Information on an Amorphous Computer. In: *AI Mcmo 1666*, MIT AI Laboratory (1999)
9. Nagpal, R., Shrobe, H., Bachrach, J.: Organizing a Global Coordinate System from Local Information on an Ad-Hoc Sensor Networks. In: Zhao, F., Guibas, L.J. (eds.) *IPSN 2003*. LNCS, vol. 2634, pp. 333–348. Springer, Heidelberg (2003)
10. He, T., Huang, C.D., Blum, B.M.: Range-Free Localization Schemes in Large Scale Sensor Networks. In: *Proceedings, 9th Annual International Conference on Mobile Computing and Networking*, San Diego, pp. 81–95 (2003)

11. Vijay, C., Winston, K.G.S., Yoo, S.C.: Localization in Underwater Sensor Networks-Survey and Challenges. In: WUWNet 2006, California, pp. 33–40 (2006)
12. Bergamo, P., Mazzini, G.: Localization in Sensor Networks with Fading and Mobility. In: 13th IEEE Int'l Symp. on Persona, Indoor and Mobile Radio Communications, Lisbon, pp. 750–754 (2002)
13. Savarese, C., Rabay, J., Langendoen, K.: Robust Positioning Algorithm for Distributed Ad-hoc Wireless Sensor Networks. In: USENIX Technical Annual Conference, Monterey, pp. 317–327 (2002)
14. Wang, F.B., Shi, L., Ren, F.Y.: Self-Localization System and Algorithms for Wireless Sensor Networks. *Journal of Software* 16, 857–866 (2006)

A Hybrid Neural Network Method for UAV Attack Route Integrated Planning

Nan Wang, Xueqiang Gu, Jing Chen, Lincheng Shen, and Min Ren

Mechatronics and Automation School of National University of Defense Technology,
410073, Changsha, China
xlaser2003@yahoo.com.cn

Abstract. This paper proposes a hybrid neural network method to solve the UAV attack route planning problem considering multiple factors. In this method, the planning procedure is decomposed by two planners: penetration planner and attack planner. The attack planner determines a candidate solution set, which adopts Gaussian Radial Basis Function Neural Networks (RBFNN) to give a quick performance evaluation to find the optimal candidate solutions. The penetration planner adopts an alternative Hopfield Neural Network (NN) to refine the candidates in a fast speed. The combined effort of the two neural networks efficiently relaxes the coupling in the planning procedure and is able to generate a near-optimal solution within low computation time. The algorithms are simple and can easily be accelerated by parallelization techniques. Detailed experiments and results are reported and analyzed.

Keywords: UAV, air-ground attack, route planning, multiple factors, RBFNN, alternative Hopfield NN.

1 Introduction

With the development of navigation and weapons technology, UAVs are playing an increasingly important role in military operations and have been used to perform a variety of high-risk tasks, such as reconnaissance, well-protected targets attack, and so on. In performing air-ground attack tasks, UAV would fly to the corresponding targets, enter the target area, and use on-board sensors to detect and track the target, and attack at an appropriate distance, while avoiding the detection and engagement of ground threats. So the UAV attack route planning should consider multiple factors, such as terrain, threat, flying constrains, sensor and weapon employment constraints.

Route Planning is a fundamental problem and extensive research efforts have been directed towards this problem [1]. To improve the efficiency of route planning methods, a lot work has been developed. Visibility graph [2] and Voronoi diagram [3] have an excellent speed in 2-D environment and can easily find the most survivable route in environments full of threats. The sparse A* search (SAS) method [4] is proposed for tactical aircraft real time route planning, which uses heuristic knowledge to efficiently increase the planning speed. Probabilistic map [5] follows a probability scheme to find routes and a coarse route can be found quickly in a complex environment. Genetic algorithm [6] uses DNA schemes to search and evaluated routes in a self-organized

way, which is useful in most route planning problems without the requirement of any prior knowledge. Neural networks [7] and potential field [8] methods, which rapidly construct a reactive layer or an artificial force field following by the route, are suitable for a majority of route planning problems, too. However, most of the reported work in route planning concentrates on target orientating and collision avoidance, and fall short in dealing with multiple factors such as sensor and weapon employment objects and constrains.

The main emphasis of this paper is the research done towards the development of a hybrid neural network based method for the UAV attack route integrated planning. Detailed algorithms are also presented.

2 Problem Specification

The UAV attack route planning is influenced by multiple factors: ①terrain; ②threats; ③target characteristics; ④flying constraints; ⑤sensor employment constraints and objects; ⑥weapons employment constraints. During penetration, the UAV flies at a constant average height H_p above ground to the target, while avoiding threats. When the attack begins, the UAV flies up at the tracking point with attack height and course ψ , finds and locks the target and delivers the weapons, considering target characteristics and meeting the sensor and weapon constraints. So the configuration space (C-space) of the problem is represented by 2-D cells during penetration and by the tracking point and attack course during attack.

The factors and their influences are modeled as follows:

- *Terrain*. Terrain influences flight safety, threat envelope, and the utilities of sensor and weapon mainly in two ways: ①Obstacle effect. Terrain over a certain height of H_{Cmax} is considered as forbidden terrain, the corresponding cell is considered as C-space obstacle. ②Terrain mask effect. There exist some invisible areas at a certain height around the ground threat and target, as shown in the figure 1. Flying through the threat masking cell is considered to be safe, while flying through the target masking cell during attack is not allowed because it may break the lock on the target. This article uses a fast algorithm to calculate terrain masking. Details of the algorithm are described in [9].

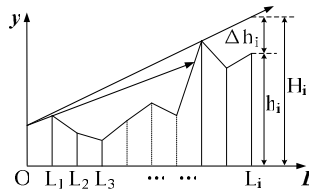


Fig. 1. The terrain mask effect is showed by Δh_i at sample point L_i from the threat or target center O. Here h_i is the height of L_i , H_i is the mask height of L_i .

- *Threats.* The general threat model is adopted to regard the non-masked threat cells as obstacle cells during penetration. When the attack begins, the threat cost K is calculated by (1), where T_s is the number of steps expose to threat, L_s is the total steps taken by the attack phase.

$$K = T_s / \text{Max}(L_s) \tag{1}$$

- *Target Characteristics.* Target characteristics contains coordinate (x_T, y_T) , height h_T , size and optimal-attack direction ψ_T . The maximum discovery range R_{SMAX} and minimum discovery range R_{SMIN} are determined by s_T .
- *Flying Constraints.* The flying constraints of the UAV are represented by not allowing the adjacent cell angle to exceed 45° in the C-space, i.e., UAV could only move forward towards three adjacent cells in every step.
- *Sensor employment constraints and objects.* The UAV can only find and locate targets in sensor-in-zone, that is, in azimuth angle θ_s , pitching angle θ_p between maximum range R_{SMAX} and minimum range R_{SMIN} ahead of its nose. Besides, the UAV must maintain line-of-sight to the target and a constant course for the stable tracking of the target. All of the above are called the sensor employment constraints C_s , in which θ_s , θ_p , R_{SMAX} and R_{SMIN} are determined by the performance of the sensor. The tracking performance E_s is calculated by (2), where w_i are weights, es_1 is the optimality of selected attack direction, es_2 is the optimality of sensor azimuth deviation, es_3 is the optimality of weapon delivery distance, es_4 is the optimality of threat exposure.

$$E_s = \begin{cases} \sum_{i=1}^4 w_i \times es_i, & C_s \leq 0 \\ 0, & \text{else} \end{cases} \tag{2}$$

- *Weapons employment constraints.* The weapons carried by UAV are usually "fire-and-forget" type. Such weapons should be delivered under certain angle and height conditions, which are modeled as relative azimuth θ_w , delivery height range $[H_{\text{MIN}}, H_{\text{MAX}}]$, and ground range $[R_{\text{WMIN}}, R_{\text{WMAX}}]$. Before weapon delivery, the UAV needs to stably track the target for distance d_s to accurately locate the target. The weapon-in-zone and d_s together are called weapon employment constrains C_w , the weapon is delivered immediately after all aspects of C_w are met.

Through the above analysis, the UAV's attack route is composed of two segments: the penetration route and the attack route. The penetration route is determined by the connected cells in 2-D C-space, and the object is to optimize the route cost performance E_r , which is calculated by route steps r and the minimum steps (straight line steps) between the start and end cell, as shown in (3). The attack route is a straight route determined by the tracking point azimuth angle $\acute{\alpha}$ and pitching angle β to the target, tracking distance R and attack direction ψ , and the object is to maximize the E_s . The two segments are joint at the tracking point, with a combined performance E calculated by (4), where w_s and w_r are weights.

$$E_r = \begin{cases} \min(r) / r, & \text{the end cell is reachable} \\ 0, & \text{else} \end{cases} \tag{3}$$

$$E = w_s \times E_s + w_r \times E_r \tag{4}$$

3 Methodology

As described in the above section, the combine configuration space for the UAV attack route integration planning can become very large, since the planner should not only determine the tracking point characteristics but also plan a route to it from the entry point. Besides, multiple factors should be considered in the planning procedure, which makes it difficult to establish and use any heuristic knowledge. As a result, conventional route planning methods fall short in dealing with it efficiently. A better way to deal with this problem is to decompose it into smaller problems as penetration route planning and attack route planning and an independent planner is applied to each problem. The attack planner determines the optimal tracking point characteristics (α , β , R , Ψ) while meeting the sensor and weapon employment constrains. The penetration planner plans the optimal route from the entry point to the tracking point. However, the two planners are coupled, since the attack planner seeks to get the best attack direction and position while the penetration phase seeks to attain a shortest path avoiding the ground threats, which may often lead to poor or even infeasible solutions. Moreover, the planning procedure would repeatedly iterate between the two planners that greatly increases the planning time, thus debases the planning efficiency.

To further deal with the above problem, a hybrid neural network method is proposed by us. In this method, two kinds of neural networks are applied to each planner, as shown in figure 2. First, for the attack planner, Radial Basis Function Neural Networks (RBFNN) are designed to search the planning space and determined a set of candidate tracking point characteristics, relying on the fast response of the network. Then for the penetration planner, an alternative Hopfield Neural Network (NN) is introduced to quickly plan the optimal routes from the entry point to the candidate tracking points that make up the synthesized candidate solution. After that, the E_R and E_S of each candidate solutions are calculated and the one with the largest E is selected as the final solution.

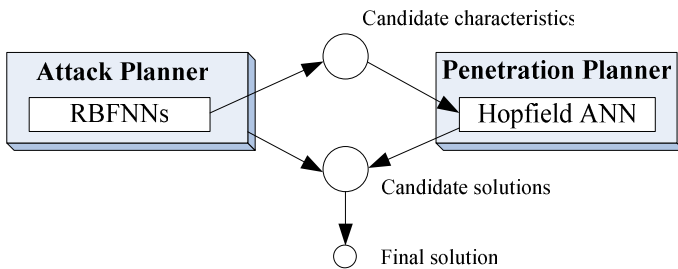


Fig. 2. The structure of the hybrid neural network method

3.1 Alternative HNN

Hopfield NN is a potential based neural network and is suitable for route planning in 2-D C-space. In this paper, an improved Hopfield NN algorithm is introduced to the penetration planner. The algorithm quickly sets up a numerical potential field from the end cell to the edge of the C-space, using distance transformation based serial simulation [10]. Then the optimal route could be found following the field gradient using

steepest ascent principle. The algorithm need not to be trained and can attain a very fast planning speed.

The planning procedure of the algorithm is as follows. First the planning space is expressed by the Hopfield network, i.e., each cell in C-space is regarded as a neuron. Each free neuron (non-obstacle neuron) can be connected with n adjacent neurons (In this the paper, n = 8), as shown in figure 3. Then, the field construction starts from the end neuron that consequently updates the output value of adjacent neurons by (5)(6) through the distance transformation based serial simulation, thus to establish the numerical potential field, where X^i is the output of the ith neuron, I is the input value, D_i is 0 for obstacle neurons and 1 for others, repnum is the iteration number, NE_i is the set of adjacent neurons of i, $A > 2n$. Finally, based on the field gradient, the optimal route from the start cell to the end cell could be found quickly using steepest ascent principle.

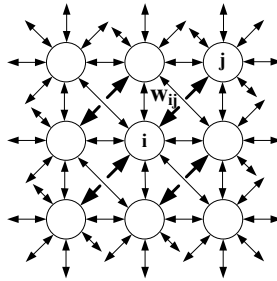


Fig. 3. The planning space expressed by the Hopfield NN, where w_{ij} is the connection weights from neuron i to j

$$X^i(repnum + 1) = \begin{cases} D_i \times f(X) / A, & i \text{ is not the end neuron} \\ D_i \times f(X) / A + I / A, & i \text{ is the end neuron} \end{cases} \quad (5)$$

$$f(X) = \sum_{j < i \cap j \in NE_i} \omega_j X^j(repnum + 1) + \sum_{j > i \cap j \in NE_i} \omega_j X^j(repnum) \quad (6)$$

3.2 Design of RBFNN

RBFNN is a universal approximator network with compact structure, which can approach to a variety of non-linear functions with fast learning speed. Here the designed RBFNN is a general RBFNN containing three layers: input layer, output layer and hidden layer, as shown in figure 4. In the hidden layer, the radial basis function is selected as Gauss function. The output y is calculated from the input vector x by (7), where t_i stands for the ith hidden neuron, w_i is the connection weight values of the ith neuron, σ is the radius of Gauss function, $\| \bullet \|$ is the Euclid norm.

$$y = \sum w_i \times \exp\left(-\|x - t_i\|^2 / 2\sigma^2\right) \quad (7)$$

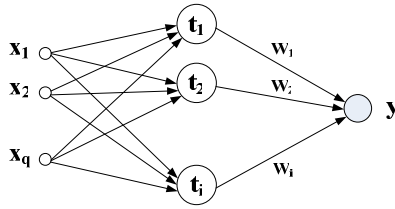


Fig. 4. RBFNN structure

In this paper, the RBFNNs are used to approximate the solution performance of the two planners. The planning space is divided into 12 zones, and a RBFNN is built in each zone to quickly response to the performance of the tested solutions. With the help of RBFNNs, the attack planner could quickly search the planning space and determine the candidate characteristics set, without repeatedly runs the penetration route planning algorithm to attain E_R , thus relaxes the coupling problem. The input vector of the RBFNN is composed of tracking point characteristics. The output vector is composed of the penetration route performance E_R and the attack route performance E_S . The hidden neurons are fixed and uniformly distributed in the sensor-in-zone around the target. Before use, the weight values of the RBFNN should be trained first.

3.3 Hybrid Neural Network Method

The planning procedure of the hybrid neural network method is as follows:

① Training of RBFNN. First, the number of the training samples are determined. And the training samples are uniformly selected in the sensor-in-zone which compose the training set. Then each sample is passed to the penetration planner and the attack planner. In the attack planner the constrains are examined and the E_S is calculated. While in the penetration planner, the optimal route from the start point to the tracking point is planned by the alternative Hopfield NN and the E_R is calculated. After the whole training set is calculated, the RBFNN is trained and the weights are updated.

②Attack route planning. With the fast response of the trained RBFNN, the tracking point characteristics(α, β, R, ψ) in the sensor-in-zone are searched and sorted by E^\wedge , which is calculated by (8), where E^\wedge_S, E^\wedge_R are the performances calculated by RBFNN.

$$\hat{E} = w_S \times \hat{E}_S + w_R \times \hat{E}_R \tag{8}$$

③Candidate characteristics selection. For the first time, let $0 \rightarrow N_0, 30 \rightarrow N_1$, the above results sorted in $[N_0, N_1]$ are selected as the candidate tracking point characteristics.

④Refined planning. For each candidate, the related penetration route is planned by the penetration planner using the alternative Hopfield NN and its true E_R is calculated, while the true E_S is calculated by the attack planner. Then the E of each candidate is updated and the candidates are re-sorted by E . If all E equals 0, then $N_{0+1} \rightarrow N_0, N_1+30 \rightarrow N_1$, back to ③; Otherwise, the one with the largest E is selected as the final integrated route plan.

4 Experiments and Results

The hybrid neural network method is tested in two different scenarios, as shown in figure 5. The simulated parameters of the factors are as follows:

Table 1. The simulated parameters of the factors

Parameters	Values	Parameters	Values	Parameters	Values
θ_S	$\pm 40^\circ$	θ_W	$\pm 60^\circ$	R_{WMIN}	1000m
θ_P	40°	H_{MAX}	5000m	d_S	2000m
R_{SMAX}	10000m	H_{MIN}	100m	ψ_T	90°
R_{SMIN}	1000m	R_{WMAX}	5000m	H_P	500m

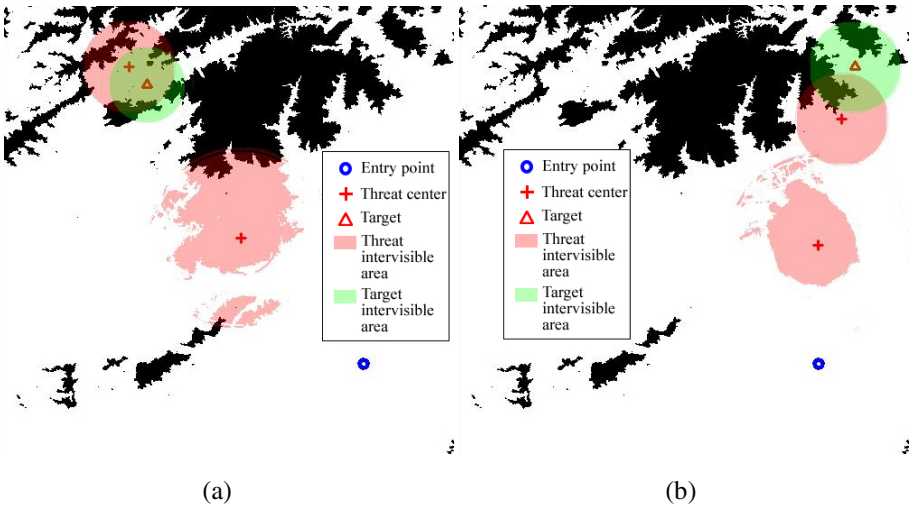


Fig. 5. Typical scenarios of the UAV air-ground attack tasks. The terrain mask is calculated and is displayed by an average attack height of 1000m near the target and H_P in other areas. The areas in black are the forbidden terrain.

The resolution of the training samples (composed of tracking point characteristics) of the RBFNN is shown in table 2.

Table 2. The distribution and resolution of the training samples

Dimensions	Contents	Resolution	Scales
1	azimuth angle α	360/12	[0, 360)
2	pitching angle β	$\theta_P / 4$	(0, θ_P]
3	tracking distance R	$(R_{SMAX} - R_{SMIN}) / 4$	$[R_{SMIN}, R_{SMAX}]$
4	attack direction ψ	$\theta_S / 2$	$[\alpha - \theta_S, \beta + \theta_S]$

The total number of the training samples is 1200, which is divided into 12 zones with 100 samples in each. So the number of the input training samples of each RBFNN is 100. The RBF centers are selected the same as the training samples so the number of hidden neurons is 100, too. After training, the RBFNN is used to search the planning space in a higher resolution and 64800 solutions are tested and compared. All the algorithms are programmed in VC++ and runs in a standard PC with Pentium 4 CPU of 2.8GHz.

The results of the experiments are as follows. Figure 6 shows the potential field built by the alternative Hopfield NN in each scenario. Figure 7 and 8 shows the optimal integrated attack routes found. Figure 9 gives a compare of the true performance E of the candidate solutions and the E^{\wedge} calculated by RBFNN. Table 3 shows the average time consumption of the algorithms.

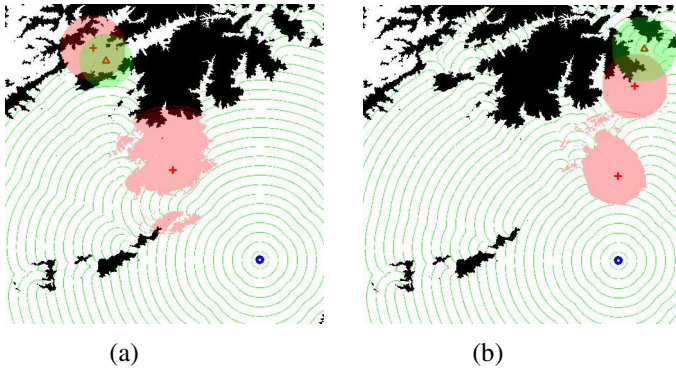


Fig. 6. The potential field built by the alternative Hopfield NN in the two scenarios

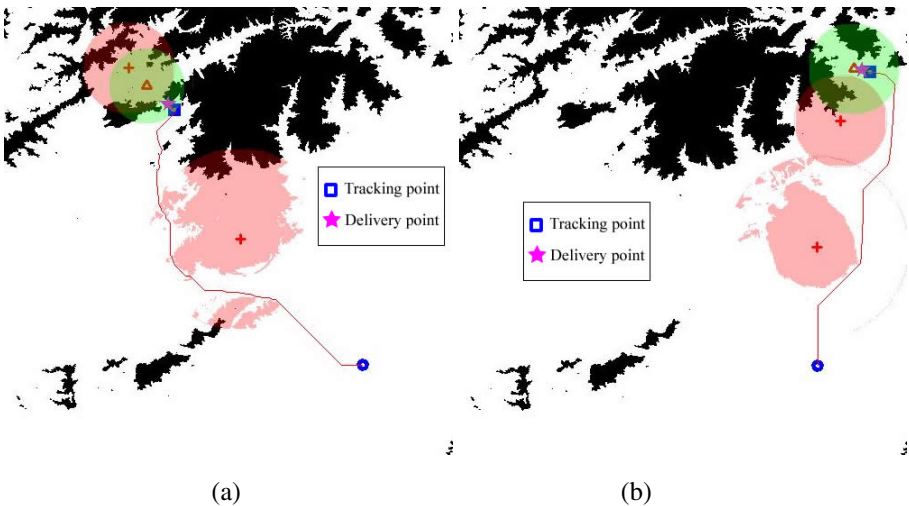


Fig. 7. The optimal integrated attack routes found in the two scenarios

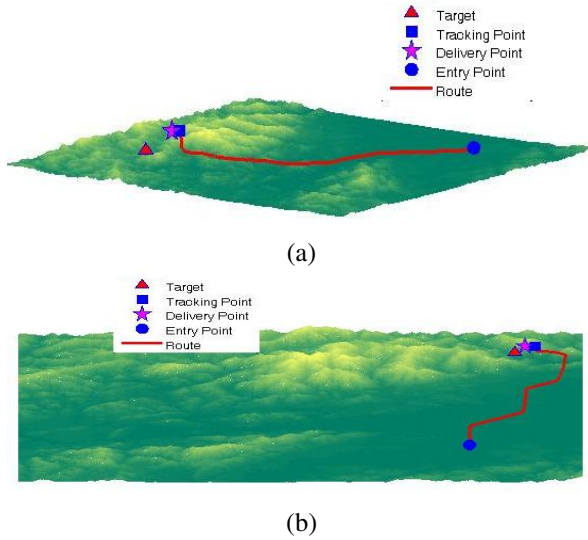


Fig. 8. 3-D display of the optimal attack route found in the two scenarios

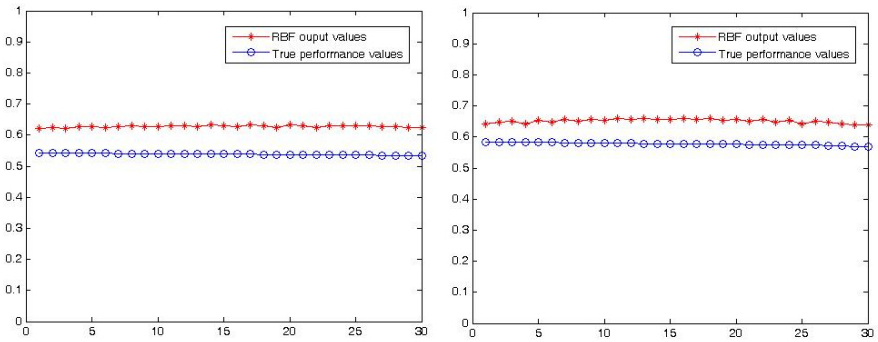


Fig. 9. True performance E of the candidate solutions V s E^A calculated by RBFNN

Table 3. Average time consumption of the algorithms

	Scenario a	Scenario b
RBFNN training time	3.617s	3.591s
RBFNN search time (64800 solutions searched)	6.616s	6.587s
Hopfield NN search time (per route)	0.019s	0.013s

5 Conclusion

A hybrid neural network based method dealing with the UAV attack route planning problem considering multiple factors is presented. In this method, the complicated

planning problem is decomposed and solved respectively by two planners: penetration planner and attack planner. To deal with the coupling problems, hybrid neural networks are adopted. One is RBFNN used in the attack planner to determine the initial candidate solutions. The other is an alternative Hopfield NN used in the penetration planner for refined planning. The combined effort of the two neural networks efficiently relaxes the coupling in the planning procedure, and makes it possible to generate a near-optimal route plan within low computation time. The algorithms are simple and can easily be accelerated by parallelization techniques. Experimental results show that this method works well in solving the UAV attack route integrated planning problems considering multiple factors.

References

1. Latombe, J.C.: Robot Motion Planning. Kluwer Academic Publishers, Boston (1991)
2. Asano, T., Asano, T., Guibas, L., Hershberger, J., Imai, H.: Visibility-polygon search and Euclidean shortest path. In: 26th Symp. Found. Comp. Science, pp. 155–164 (1985)
3. Bortoff, S.A.: Path Planning for UAVs. In: Proceedings of the 2000 American Control Conference, vol. 1, pp. 364–368 (2000)
4. Robot, J.S., Peggy, G., Ira, S.G.: Robust algorithm for real-time route planning. IEEE Transactions on Aerospace and Electronic Systems 36, 869–878 (2000)
5. Overmars, M.H., Svestka, P.: A probabilistic learning approach to motion planning. J. In: Proc. Workshop Algorithmic Foundations Robotics, pp. 19–37 (1994)
6. Solano, J., Jones, D.I.: Generation of collision-free paths, a genetic approach. IEEE Colloquium on Gen. Alg. for Control Sys. Eng., 5/1–5/6 (1993)
7. Kozakiewicz, C., Ejiri, M.: Neural network approach to path planning for two dimensional robot motion. Proc. IEEE/RSJ (IROS 1991) 2, 818–823 (1991)
8. Khatib, O.: Real-Time Obstacle Avoidance for Manipulators and Mobile Robots. Int. J. of Robotics Research 5(1), 90–99 (1986)
9. Leavitt, C.A.: Real-Time In-Flight Planning. Proceedings of the IEEE 1996 National Conference on Aerospace and Electronics 1, 83–89 (1996)
10. Fan, C., Lu, Y., Liu, H., Huang, S.: Path Planning for Mobile Robot Based on Neural Networks (in Chinese). Computer Engineering and Applications 8, 86–89 (2004)
11. Rana, A.S., Zalzala, A.M.S.: A Neural Networks Based Collision Detection Engine for Multi-Arm Robotic Systems. In: 5th International conference on artificial neural networks, pp. 140–145 (1997)
12. Ellips, M., Davoud, S.: Classic and Heuristic Approaches in Robot Motion Planning—A Chronological Review. Proceedings of world academy of science, engineering and technology 23, 101–106 (2007)
13. Zou, A., Hou, Z., Fu, S., Tan, M.: Neural Networks for Mobile Robot Navigation: A Survey. In: Wang, J., Yi, Z., Žurada, J.M., Lu, B.-L., Yin, H. (eds.) ISNN 2006. LNCS, vol. 3973, pp. 1218–1226. Springer, Heidelberg (2006)
14. Pashkevich, A., Kazheunikau, M.: Neural network approach to trajectory synthesis for robotic manipulators. Journal of Intelligent Manufacturing 16, 173–187 (2005)
15. Ranka, K., Zoran, V.: Methodology of Concept Control Synthesis to Avoid Unmoving and Moving Obstacles. Journal of Intelligent and Robotic Systems 45, 267–294 (2006)
16. Guang, Y., Vikram, K.: Optimal path planning for unmanned air vehicles with kinematic and tactical constraints. In: Proceedings of the 41th IEEE Conference on Decision and Control, vol. 2, pp. 1301–1306 (2002)

Hybrid Game Theory and D-S Evidence Approach to Multiple UCAVs Cooperative Air Combat Decision

Xingxing Wei, Haibin Duan, and Yanran Wang

School of Automation Science and Electrical Engineering, Beihang University,
Beijing 100191, P R China
starswei@gmail.com, hbduan@buaa.edu.cn, yanran22kk@126.com

Abstract. Mission decision-making is one of the most important techniques for cooperative combat of multiple unmanned combat aerial vehicles (UCAVs), while game theory is an efficient method in solving mission decision-making. In this paper, the game theory is applied to the air combat decision of multiple UCAVs. Then a weapon model for UCAVs' air-to-air missiles is proposed to obtain the basic probability value of the D-S evidence theory. In order to obtain the Nash equilibrium point, the bimatrix problem is transferred into an optimization problem. In this way, the function "linprog" in MATLAB can be used to obtain the optimal solution, which can greatly simplify the steps of solving the bimatrix problem. Finally, the optimal strategy is obtained by optimizing calculation. Series of experimental results demonstrate the feasibility and effectiveness of the proposed approach in solving the multiple UCAVs cooperative air combat decision problem.

Keywords: UCAVs, Game Theory, Mission Decision-making, D-S Evidence Theory.

1 Introduction

Following with the development of unmanned combat aerial vehicles' (UCAV) technology, cooperative unmanned aerial vehicles control will receive a great attention in the future. For many military missions, it is unthinkable to be taken on a single UCAV. And more instances show that it is more complicated and expensive to exploit a single UCAV than to create a multiple UCAVs' system. Wang, H.L. analyzed a maneuvering-decision method for air-to-air combat [1], Yang, Y.C. proposed quantitative air combat decision approach in [2], and Wang, Y.N presented an intelligent differential game on air combat decision [3]. Yao, Z.X. also presented the game theory model to solve this problem, and tests prove that it can conduct the mission decision-making effectively [4]. But these approaches to establish the basic probability value and solve the bimatrix are very complicated.

To overcome the above-mentioned shortcomings, we present a weapon model of UCAVs' air-to-air missiles to obtain the basic probability value. In order to solve the bimatrix, we transfer it into an optimization problem first, and then the function "linprog" in MATLAB is used to solve the optimization problem. Furthermore, we apply the method on the issue of multiple UCAVs air combat, and design corresponding

strategy sets. Finally, series of experimental results are given to demonstrate the feasibility and effectiveness of the proposed approach in solving the multiple UCAVs cooperative air combat decision problems.

2 Introduction to Related Theories

2.1 D-S Evidence Theory

D-S evidence theory is proposed by Dempster in 1976, then his student Shafer develops and organizes it into a comprehensive mathematical theory. Through synthesizing several evidences, D-S evidence theory can improve the dependable degree of the proposition, now the theory has specifically applied on some areas, such as multiple sensors network, medical diagnosis and so on. In D-S evidence theory, the combinatorial formula is very important [5]. If θ is the discriminating frame, 2^θ is the aggregate composed by all its subsets, m_1, m_2 are the basic probability values of the independent authentic function of 2^θ , A_i and B_j are the focus respectively, then the D-S combinatorial formula is:

$$m(C) = \begin{cases} 0, & C = \phi \\ \frac{\sum_{A_i \cap B_j = C} m_1(A_i)m_2(B_j), \forall C \subset \theta, C \neq \phi}{1 - K_1} \end{cases} \tag{1}$$

Where Φ is the empty aggregate, $C = A_i \cap B_j \neq \Phi$ is the focus to be fused.

$$K_1 = \sum_{A_i \cap B_j = \Phi} m_1(A_i)m_2(B_j) < 1 \tag{2}$$

2.2 Game Theory

Game theory, defined mathematically by Nash J.F [11], has found its first applications in economics, especially to solve the problems concerning the decisions that have some effects on different and often competitive fields.

In a model of game theory, three elements must exist. They are players, strategy set, and criterion. Each player has its own strategy set and its own criterion. When a game begins, each player searches its own best strategy in its search space to improve its own criterion with all the rest criteria fixed by others players. So there exists the exchange of strategies among the players. The frequency of exchanged δ is called the Nash frequency, generally $\delta=1$, which means the exchange of best strategies happens at the end of each generation. When no player can further improve its criterion, it means that the system has reached a state of equilibrium called Nash equilibrium[6].

We take a two-player game to present the process of Nash equilibrium.

Let A be the search space for the first player, B the search space for the second player, a strategy pair $(x^*, y^*) \in A \cdot B$ is said to be a Nash equilibrium iff:

$$\begin{aligned} f_A(x^*, y^*) &= \inf_{x \in A} f_A(x, y^*) \\ f_B(x^*, y^*) &= \inf_{x \in B} f_B(x, y^*) \end{aligned} \tag{3}$$

Where f_A is the gain for the first player, f_B is the gain for the second player.

3 Mission Decision-Making Method

3.1 A Mission Scenario

The red UCAV formation attacks the blue ground target, its main task is to attack the blue airport (G_1). The formation includes an unmanned fighter-bomber (R_1) and $m-1$ UCAVs (R_2, R_3, \dots, R_m). The unmanned fighter-bomber's mission is to bomb the blue airport and the other UCAVs are to help it complete the bombing mission. $m-1$ UCAVs are exactly the same, i.e. they have the same flight performance and weapon performance. Each UCAV carries several air-to-air missiles and they are their main tools for fighting. Air-to-air missile's shooting average is determined by q situation factors (such as distance, angle and speed and so on) in a battlefield situation. The unmanned fighter-bomber's maneuverability is lower. Except for air-to-air missiles it carries air-to-surface missiles to bomb the blue airport yet. (In the following paper, we don't distinguish the unmanned fighter-bomber and the $m-1$ UCAVs and all are named as UCAVs).

The blue k UCAVs (B_1, B_2, \dots, B_k) composed an attack formation. Their mission is to prevent the fighter-bomber from bombing the airport and annihilate them. The UCAVs in the formation are also exactly the same, i.e. each UCAV carries air-to-air missiles and they are their main tools for fighting. Air-to-air missile's shooting average is determined by q situation factors (such as distance, angle and speed and so on) in a battlefield situation.

The blue strategy set $S_1 = (\alpha_1, \alpha_2, \dots, \alpha_w)$ includes w strategies. The aggregate composed by the blue k UCAVs is defined as WU . The aggregate composed by all subsets of the WU is 2^{WU} , so in the blue strategy set, the strategy α_i :

$$\alpha_i = ((T_{i1}, R_1), (T_{i2}, R_2), \dots, (T_{il}, R_l))$$

$$T_{ib} = 2^{WU}, \quad b = 1, 2, \dots, k; \bigcup_{k=1}^l T_{ik} = WU; T_{ig} \cap T_{ih} = \emptyset, \forall g, h, g = 1, 2, \dots, l, h = 1, 2, \dots, l, g \neq h.$$

denotes the blue puts the UCAVs in aggregate T_{i1} attack the target R_1 cooperatively, puts the UCAVs in aggregate T_{i2} attack the target R_2 cooperatively and so on.

The red strategy set $S_2 = (\beta_1, \beta_2, \dots, \beta_d)$ includes d strategies. The aggregate composed by the red m UCAVs is defined as DU . The aggregate composed by all subsets of the DU is 2^{DU} , so in the red strategy set, the strategy β_i :

$$\beta_i = ((T_{i1}, B_1), (T_{i2}, B_2), \dots, (T_{ik}, B_k))$$

$$T_{ip} = 2^{DU}, \quad b = 1, 2, \dots, k; \bigcup_{p=1}^k T_{ip} = DU; T_{ig} \cap T_{ih} = \emptyset, \forall g, h, g = 1, 2, \dots, k, h = 1, 2, \dots, k, g \neq h.$$

denotes the red puts the UCAVs in aggregate T_{i1} attack the target B_1 cooperatively, puts the UCAVs in aggregate T_{i2} attack the target B_2 cooperatively and so on.

Therefore the model of game theory in this paper is: $G = \langle N, S_1, S_2, u_1, u_2 \rangle$, $N = \{1, 2\}$, 1 denotes the blue formation (including k UCAVs), 2 denotes the red formation (including an unmanned fighter-bomber and $m-1$ UCAVs). S_1 is the blue strategy set, S_2 is the red strategy set. u_1 is the blue payoff function, u_2 is the red payoff function.

3.2 Evidence Syntheses

D-S evidence theory uses the synthesis of multiple evidence to make decision. It has a great deal of flexibility in dealing with the unknown and uncertainty. In this paper, the high-level evidence is obtained through the synthesis of low-level evidence until the effectiveness (income) and the invalidity (price) of some attack (or defense) strategy are obtained.

The D-S basal combinatorial formula of the mission decision-making method is below:

$$m1^{\alpha i_{nt}} = m1^{\alpha i_{1nt}} \oplus m1^{\alpha i_{2nt}} \oplus \dots \oplus m1^{\alpha i_{qnt}} \tag{4}$$

Where , $q = 1,2,\dots,q; n = 1,2,\dots,n; t = 1,2,\dots,t;$

denotes the basic probability value when the n -thUCAV in blue formation attacks the t -thUCAV in red formation considering q momentum factors (in this paper $q=3$, i.e. the distance between twoUCAVs, the speed of theUCAVs and the angles between theUCAVs).

The basic probability value when the blue p UCAVs attack the red t -thUCAV cooperatively can be expressed as follows:

$$m1^{\alpha i_t} = m1^{\alpha i_{1t}} \oplus m1^{\alpha i_{2t}} \oplus \dots \oplus m1^{\alpha i_{pt}} \tag{5}$$

Where , $p = 1,2,\dots,p; t = 1,2,\dots,t;$

The basic probability value when the blue selects the strategy α_i can be expressed as follows:

$$m1^{\alpha i} = m1^{\alpha i_1} \oplus m1^{\alpha i_2} \oplus \dots \oplus m1^{\alpha i_h} \tag{6}$$

The basic probability value when the blue p -thUCAV attacks the red t_1 -th, t_2 -th... t_q -thUCAV at the same time can be expressed as follows:

$$m1^{\alpha i_{p:(t_1,t_2,\dots,t_q)}} = (m1^{\alpha i_{pt_1}} + m1^{\alpha i_{pt_2}} + \dots + m1^{\alpha i_{pt_q}}) / q \tag{7}$$

By the same token, we can synthesis the red every evidence $m2^{\beta i_{nt}}$, $m2^{\beta i_t}$, $m2^{\beta i}$, $m2^{\beta i_{p:(t_1,t_2,\dots,t_q)}}$. The formula below:

$$u_1(\alpha_i \beta_j) = \frac{m1^{\alpha_i}(a) \bullet m2^{\beta_j}(b)}{m1^{\alpha_i}(b) \bullet m2^{\beta_j}(a)} \tag{8}$$

is defined as the blue payoff function when the blue selects the strategy α_i and the red selects the strategy β_j .

$$u_2(\alpha_i \beta_j) = \frac{m1^{\alpha_i}(b) \bullet m2^{\beta_j}(a)}{m1^{\alpha_i}(a) \bullet m2^{\beta_j}(b)} \tag{9}$$

is defined as the red payoff function when the blue selects the strategy α_i and the red selects the strategy β_j .

All of the above evidence has both the effectiveness and invalidity. The letters "a", "b" are used to characterize the both separately: $m1^{\alpha_i}(a)$ and $m^{\alpha}(b)$ are defined as the effectiveness and the invalidity of attack when the blue selects the strategy α_i . $m2^{\beta_j}(a)$ and $m2^{\beta_j}(b)$ are defined as the effectiveness and the invalidity of attack when the red selects the strategy β_j . Therefore, the blue payoff matrix $A = (a_{ij})_{k \times l}$ and the red payoff matrix $B = (b_{ij})_{k \times l}$ can be obtained.

3.3 The Acquisition of the Basic Probability Value

In order to get the $m1^{\alpha_i}_{1pt}$, $m1^{\alpha_i}_{2pt}$, ..., $m1^{\alpha_i}_{qnt}$, A database is established, the database can be obtained from a large number of air-to-air missile tests. Considering calculating easily, the database is put into an abstract linear.

Table 1. Distance definition in weapon model of UCAVs' air-air missiles

DISTANCE(D) /km	<=1	2	3	4	5	6	7	8	>=9
Attack effectiveness $m^{\alpha(\beta_j)}_{1pt}(a)$	0.9500	0.8438	0.7375	0.6312	0.5250	0.4187	0.3125	0.2062	0.1000
Attack invalidity $m^{\alpha(\beta_j)}_{1pt}(b)$	0.0300	0.1162	0.2025	0.2888	0.3750	0.5012	0.6275	0.7538	0.8800
Attack uncertainty $m^{\alpha(\beta_j)}_{1pt}(\theta)$	0.0200	0.0400	0.0600	0.0800	0.1000	0.0800	0.0600	0.0400	0.0200

Table 2. Speed definition in weapon model of UCAVs' air-air missiles

SPEED(V) /m/s	<200	300	400	500	600	700	800	900	>=1000
Attack effectiveness $m^{\alpha(\beta_j)}_{2pt}(a)$	0.8900	0.7900	0.6900	0.5900	0.4900	0.3900	0.2900	0.1900	0.0900
Attack invalidity $m^{\alpha(\beta_j)}_{2pt}(b)$	0.0800	0.1575	0.2350	0.3125	0.3900	0.5125	0.6350	0.7575	0.8800
Attack uncertainty $m^{\alpha(\beta_j)}_{2pt}(\theta)$	0.0300	0.0525	0.0750	0.0975	0.1200	0.0975	0.0750	0.0525	0.0300

Table 3. Angle definition in weapon model of UCAVs' air-air missiles

ANGLE (S_{ij}) / °	180	160	140	120	100	80	60	40	20	0
Attack effectiveness $m^{\alpha(\beta_j)}_{3pt}(a)$	0.9900	0.9356	0.8811	0.8267	0.7722	0.7178	0.6633	0.6089	0.5544	0.5000
Attack invalidity $m^{\alpha(\beta_j)}_{3pt}(b)$	0.0100	0.0478	0.0856	0.1233	0.1611	0.1989	0.2367	0.2744	0.3122	0.3500
Attack uncertainty $m^{\alpha(\beta_j)}_{3pt}(\theta)$	0	0.0167	0.0333	0.0500	0.0667	0.0833	0.1000	0.1167	0.1333	0.1500

Table 3. (Continued)

ANGLE (S_{ij}) / °	-20	-40	-60	-80	-100	-120	-140	-160	-180	
Attack effectiveness $m_{\alpha_i(\beta_j)}^{3pt(a)}$	0.4456	0.3911	0.3367	0.2822	0.2278	0.1733	0.1189	0.0644	0.0100	
Attack invalidity $m_{\alpha_i(\beta_j)}^{3pt(b)}$	0.4211	0.4922	0.5633	0.6344	0.7056	0.7767	0.8478	0.9189	0.9900	
Attack uncertainty $m_{\alpha_i(\beta_j)}^{3pt(\theta)}$	0.1333	0.1167	0.1000	0.0833	0.0667	0.0500	0.0333	0.0167	0	

DISTANCE (D) is defined as the real length between the centers of two UCAVs. SPEED (V) is defined as the speed of the other side when one side calculates its basic probability value. ANGLE(S_{ij}) is defined as $S_{ij} = \alpha_j - \alpha_i$ which denotes the angle when one side calculates its basic probability value.

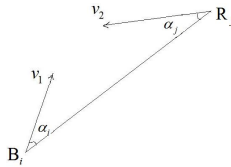


Fig. 1. The relation of angle between two UCAV ($S_{ij} = \alpha_j - \alpha_i$)

For example, when the blue i -th UCAV calculates its own basic probability value against the red j -th UCAV, the α_i in Fig.1 stands for the angle composed of the blue i -th UCAV’s velocity vector and its line of sight towards the red. The α_j stands for the angle between the red j -th UCAV’s velocity vector and line of sight towards the blue.

In addition, the basic probability value is defined when the uninhabited fighter-bomber bombs the airport as:

$$m2^{\beta_j}_{R_i G_1}(a) = 0.7345, \quad m2^{\beta_j}_{R_i G_1}(b) = 0.2277, \quad m2^{\beta_j}_{R_i G_1}(\theta) = 0.0378$$

By searching the database, the basic probability values can be obtained. If the number is not the existing numbers, it can be sought through the linear relationship. For example, if the distance between two UCAVs is d km, its corresponding basic probability values are:

$$\begin{aligned}
 m_{\alpha_i(\beta_j)}^{1pt(a)} &= \frac{0.8438 - 0.2062}{8 - 2} \cdot (8 - d) + 0.2062 \\
 m_{\alpha_i(\beta_j)}^{1pt(b)} &= 0.1 - 0.02 \times |5 - d| \\
 m_{\alpha_i(\beta_j)}^{1pt(\theta)} &= 1 - m_{\alpha_i(\beta_j)}^{1pt(a)} - m_{\alpha_i(\beta_j)}^{1pt(b)}
 \end{aligned}
 \tag{10}$$

In the similar way, the basic probability values based on speed and angle can be obtained: $(m^{\alpha_i(\beta_j)}_{2pt(a)} , m^{\alpha_i(\beta_j)}_{2pt(b)} , m^{\alpha_i(\beta_j)}_{2pt(\theta)} , m^{\alpha_i(\beta_j)}_{3pt(a)} , m^{\alpha_i(\beta_j)}_{3pt(b)} , m^{\alpha_i(\beta_j)}_{3pt(\theta)})$.

3.4 Solving the Bimatrix

The bimatrix problem can be transferred into an optimization problem [7], and the bimatrix are as follows:

$$A = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \dots & \dots & \dots & \dots \\ c_{m1} & c_{m2} & \dots & c_{mn} \end{bmatrix} \qquad B = \begin{bmatrix} \dot{c}_{11} & \dot{c}_{12} & \dots & \dot{c}_{1n} \\ \dot{c}_{21} & \dot{c}_{22} & \dots & \dot{c}_{2n} \\ \dots & \dots & \dots & \dots \\ \dot{c}_{m1} & \dot{c}_{m2} & \dots & \dot{c}_{mn} \end{bmatrix}$$

$$\begin{array}{ll}
 \min v & \min w \\
 A_i \cdot y \leq v, \quad i=1,2,\dots,n & x^T \cdot B_j \leq w, \quad j=1,2,\dots,m \\
 y_1 + y_1 + \dots + y_m = 1, & x_1 + x_1 + \dots + x_m = 1, \\
 y_j > 0, \quad j=1,2,\dots,n & x_i > 0, \quad i=1,2,\dots,m
 \end{array} \tag{11} \tag{12}$$

the formula (11) can be transferred into (11)' and the formula (12) to (12)':

$$\begin{array}{ll}
 \min v & \min w \\
 A_i \cdot y - v \leq 0, \quad i=1,2,\dots,n & x^T \cdot B_j - w \leq 0, \quad j=1,2,\dots,m \\
 y_1 + y_1 + \dots + y_m = 1, & x_1 + x_1 + \dots + x_m = 1, \\
 y_j, v > 0, \quad j=1,2,\dots,n & x_i, w > 0, \quad i=1,2,\dots,m
 \end{array} \tag{11)' \tag{12)'$$

In MATLAB, "linprog" is an efficient function to solve the above optimization problem. In this way, the Nash equilibrium point can be obtained.

If there is no optimal value, the mixed strategies needs to be solved. Assume (x^*, y^*) is the Nash equilibrium point. If $x = \max(x^*_1, x^*_2, \dots, x^*_m) = x^*_i \quad i=1,2,\dots,m$,

$$y = \max(y^*_1, y^*_2, \dots, y^*_n) = y^*_j \quad j=1,2,\dots,n$$

Where $x^* = (x^*_1, x^*_2, \dots, x^*_m)$; $y^* = (y^*_1, y^*_2, \dots, y^*_n)$; α_i (Its probability is x^*_i) is the strategy the blue should select. β_j (Its probability is y^*_j) is the strategy the red should select.

4 Simulation Experiments

In order to investigate the feasibility and effectiveness of the proposed hybrid game theory and D-S evidence approach, a series of experiments are conducted in this section.

4.1 Mission Scenario

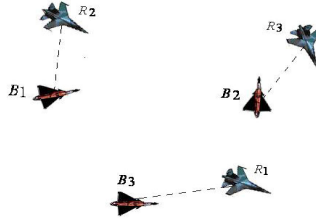


Fig. 2. Multiple UCAVs battlefield situation (R_1, R_2, R_3 belongs to the red, and B_1, B_2, B_3 belong to the blue)

Table 4. The relations of distance, angle and speed between UCAVs

	B_1-R_1	B_1-R_2	B_1-R_3	B_2-R_1	B_2-R_2	B_2-R_3	B_3-R_1	B_3-R_2	B_3-R_3
Distance (km)	8	4	10	4	11	5	5	7	7
Speed (m/s)	720	660	800	720	660	800	720	660	800
Angle (°)	90	-10	-10	-90	-80	60	90	-12	0
	R_1-B_1	R_1-B_2	R_1-B_3	R_2-B_1	R_2-B_2	R_2-B_3	R_3-B_1	R_3-B_2	R_3-B_3
Distance(km)	8	4	5	4	11	7	10	5	7
Speed (m/s)	360	660	800	360	660	800	360	660	800
Angle (°)	-90	90	90	10	80	12	10	-60	0

The blue strategy set:

The red strategy set:

$$\begin{aligned}
 \alpha_1 &= ((B_1, R_1), (B_2, R_2), (B_3, R_3)) \\
 \alpha_2 &= ((B_1, R_1), (B_2, R_3), (B_3, R_2)) \\
 \alpha_3 &= ((B_1, R_2), (B_2, R_3), (B_3, R_1)) \\
 \alpha_4 &= ((B_1, R_2), (B_2, R_1), (B_3, R_3)) \\
 \alpha_5 &= ((B_1, R_3), (B_2, R_1), (B_3, R_2)) \\
 \alpha_6 &= ((B_1, R_3), (B_2, R_2), (B_3, R_1)) \\
 \alpha_7 &= (((B_1, B_2), R_1), (B_3, (R_2, R_3))) \\
 \alpha_8 &= (((B_1, B_3), R_1), (B_2, (R_2, R_3))) \\
 \alpha_9 &= (((B_2, B_3), R_1), (B_1, (R_2, R_3)))
 \end{aligned}$$

$$\begin{aligned}
 \beta_1 &= ((R_1, B_1), (R_2, B_2), (R_3, B_3)) \\
 \beta_2 &= ((R_1, B_1), (R_2, B_3), (R_3, B_2)) \\
 \beta_3 &= ((R_1, B_2), (R_2, B_3), (R_3, B_1)) \\
 \beta_4 &= ((R_1, B_2), (R_2, B_1), (R_3, B_3)) \\
 \beta_5 &= ((R_1, B_3), (R_2, B_1), (R_3, B_2)) \\
 \beta_6 &= ((R_1, B_3), (R_2, B_2), (R_3, B_1)) \\
 \beta_7 &= ((R_1, G_1), (R_2, (B_1, B_2)), (R_3, B_3)) \\
 \beta_8 &= ((R_1, G_1), (R_2, (B_1, B_3)), (R_3, B_2)) \\
 \beta_9 &= ((R_1, G_1), (R_2, (B_2, B_3)), (R_3, B_1)) \\
 \beta_{10} &= ((R_1, G_1), (R_3, (B_1, B_2)), (R_2, B_3)) \\
 \beta_{11} &= ((R_1, G_1), (R_3, (B_1, B_3)), (R_2, B_2)) \\
 \beta_{12} &= ((R_1, G_1), (R_3, (B_2, B_3)), (R_2, B_1))
 \end{aligned}$$

4.2 Simulation Results

The blue payoff matrix:

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}	β_{11}	β_{12}
α_1	0.2603	0.1924	0.0039	0.0004	0.0005	0.0068	0.0035	0.0017	0.0047	0.0079	0.0060	0.0004
α_2	10.196	7.5372	0.1510	0.0140	0.0182	0.2653	0.1385	0.0666	0.1839	0.3095	0.2359	0.0142
α_3	395.08	292.04	5.8491	0.5426	0.7049	10.278	5.3656	2.5788	7.1263	11.993	9.1415	0.5510
α_4	9.5824	7.0832	0.1419	0.0132	0.0171	0.2493	0.1301	0.0625	0.1728	0.2909	0.2217	0.0134
α_5	0.6498	0.4803	0.0096	0.0009	0.0012	0.0169	0.0088	0.0042	0.0117	0.0197	0.0150	0.0009
α_6	0.6841	0.5057	0.0101	0.0009	0.0012	0.0178	0.0093	0.0045	0.0123	0.0208	0.0158	0.0010
α_7	2.7147	2.0067	0.0402	0.0037	0.0048	0.0706	0.0369	0.0177	0.0490	0.0824	0.0628	0.0038
α_8	30.400	22.471	0.4501	0.0418	0.0542	0.7909	0.4129	0.1984	0.5483	0.9229	0.7034	0.0424
α_9	8.6004	6.3573	0.1273	0.0118	0.0153	0.2237	0.1168	0.0561	0.1551	0.2611	0.1990	0.0120

The red payoff matrix:

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}	β_{11}	β_{12}
α_1	3.8411	5.1963	259.45	2796.5	2152.9	147.65	282.82	588.46	212.95	126.53	166.00	2753.9
α_2	0.0981	0.1327	6.6243	71.402	54.970	3.7698	7.2212	15.025	5.4371	3.2306	4.2385	70.314
α_3	0.0025	0.0034	0.1710	1.8428	1.4187	0.0973	0.1864	0.3878	0.1403	0.0834	0.1094	1.8147
α_4	0.1044	0.1412	7.0489	75.978	58.494	4.0114	7.6840	15.988	5.7856	3.4377	4.5101	74.821
α_5	1.5389	2.0818	103.94	1120.4	862.56	59.153	113.31	235.76	85.314	50.692	66.507	1103.3
α_6	1.4618	1.9776	98.741	1064.3	819.39	56.192	107.64	223.96	81.045	48.155	63.178	1048.1
α_7	0.3684	0.4983	24.881	268.19	206.47	14.159	27.123	56.434	20.422	12.134	15.920	264.10
α_8	0.0329	0.0445	2.2219	23.949	18.438	1.2644	2.4221	5.0395	1.8237	1.0836	1.4216	23.584
α_9	0.1163	0.1573	7.8537	84.653	65.173	4.4694	8.5614	17.813	6.4462	3.8302	5.0251	83.364

By calculating, the conclusion can be obtained :

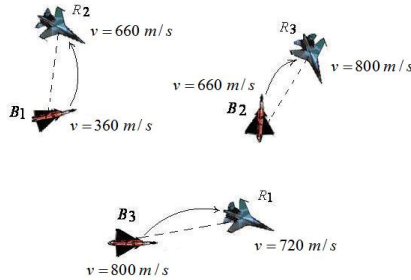


Fig. 3. The blue target allocation (B_1 against R_2 , B_2 against R_3 , B_3 against R_1)

The blue should select the strategy $\alpha_3 = ((B_1, R_2), (B_2, R_3), (B_3, R_1))$ and the value of its payment is 0.5426. The red should select the strategy $\beta_4 = ((R_2, B_1), (R_1, B_2), (R_3, B_3))$ and the value of its payment is 1.843. The blue and red target allocation results can be shown with Fig 3 and Fig 4. It is obvious that the target allocation is reasonable. In reality, they are corresponding with the actual situation. The simulation results demonstrate the proposed approach to multiple UCAVs cooperative air combat decision is feasible and effective.

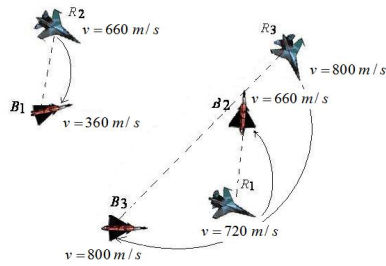


Fig. 4. The red target allocation (B_1 against R_2 , B_2 against R_1 , B_3 against R_3)

5 Conclusions

This paper has proposed hybrid game theory and D-S evidence approach to multiple UCAVs cooperative air combat decision. Series of experimental results demonstrate the feasibility and effectiveness of the proposed approach in solving the multiple UCAVs cooperative air combat decision problems. Our future work will focus on applying the proposed hybrid method to real multiple UCAVs cooperative air combat in more complicated combating environments.

Acknowledgments. This work was supported by Natural Science Foundation of China under grant #60604009, Aeronautical Science Foundation of China under grant #2006ZC51039, “Beijing NOVA Program” Foundation of China under grant #2007A0017, “New Star in Blue Sky” talent program of Beihang University, and “Science Research Training Program” of Beihang University.

References

1. Wang, H.L., Tong, M.G.: Maneuvering-Decision Analysis for Air-To-Air Combat. *Acta Aeronauticae Astronautica Sinica* 18, 371–374 (1997)
2. Yang, Y.C., Jiang, Y.X.: Quantitative Air Combat Decision. *Journal of Beijing University of Aeronautics and Astronautics* 31, 869–873 (2005)
3. Wang, Y.N., Jiang, Y.X.: An Intelligent Differential Game on Air Combat Decision. *Flying Dynamics* 21, 66–70 (2003)
4. Yao, Z.X., Li, M., Chen, Z.J.: Mission Decision-making Method of Multi-aircraft Cooperative Attack Multi-object Based on Game Theory Model. *Aeronautical Computing Technique* 37, 7–11 (2007)
5. Wang, B., Liang, G.Q., Wang, C.L.: D-S Algorithm Based on Particle Swarm Optimizer. In: 8th International Conference on Electronic Measurement and Instruments, ICEMI (2007)
6. Wang, J.F., Wu, Y.Z.: Combinatorial Optimization using Genetic Algorithms and Game Theory for High Lift Configuration in Aerodynamics. In: 41th Aerospace Science Meeting and Exhibit, Reno, Nevada, pp. 6–9 (2003)

7. Zhou, X.S., Su, W.H.: A Method for Computing the Solutions of The Bimatrix Game. *Journal of Zhejiang Gongshang University* 33, 34–36 (2006)
8. Zhou, X.S.: The Simplex Method for Computing the All Solution s of theMatrix Coun termeasure. *Ma. Thema Tics in Pract Ice and Theory* 133, 42–46 (2003)
9. Yao, Z.X., Li, M., Chen, Z.J.: Simulated Research on Mission Decision-making ofUCAV Suppressing of Enemy Air Defenses. In: 1st China Navigation, Control and Guidance Conference, Beijing (2007)
10. Fu, L., Yu, M.X., Xu, X.H.: The Strategy Studying of Air Combat about the Unmanned Combat Air Vehicles. In: *Proceedings of 2008 Chinese Control and Decision Conference, Yantai* (2008)
11. Nash, J.F.: Non-cooperative Games. *Annals of Mathematics* 54(2), 286–295 (1951)

FCMAC Based Guidance Law for Lifting Reentry Vehicles

Hao Wu¹, Chuanfeng Li^{1,2}, and Yongji Wang¹

¹ Department of Control Science and Engineering, Huazhong University of Science and Technology, Luoyu Road.1037, Wuhan 430074, China

² Department of Computer and Information Engineering, Luoyang Institute of Science and Technology, Wangcheng Avenue 90, Luoyang 471023, China

sunnywu@smail.hust.edu.cn, lichuanfeng@smail.hust.edu.cn, wangyjch@mail.hust.edu.cn

Abstract. An improved nominal trajectory based guidance law for lifting reentry vehicle is proposed. In longitudinal guidance, an integrated controller comprised of LQR and FCMAC is utilized. LQR method is adopted to design state feedback controller according to the linearized models. FCMAC are introduced to correct the value of angle of attack and bank angle adaptively, by which the tracking performance of radial distance and range-to-go along the nominal trajectory is enhanced. In lateral guidance, a crossrange corridor is established to determine bank reversals according to dynamic adjusting criterion. 3DOF simulations for a lifting reentry vehicle model demonstrated the validity of this improved method.

Keywords: Lifting reentry vehicle, Reentry guidance law, LQR method, FCMAC, Nominal trajectory tracking.

1 Introduction

It is well known that the atmospheric reentry is the most critical phase of operation for reentry vehicles due to various perturbation and aerodynamic uncertainties. So reentry guidance algorithm plays an important role in steering the vehicle safely with desired requirements in large flight envelope.

The guidance design method may be classified into two main categories: nominal trajectory based technique and predictor corrector method [1]. Traditional methods like tracking nominal drag acceleration-vs-velocity profile with changed sign of bank angle command according to bank reversal corridor was widely utilized [1]. Evolved acceleration guidance logic for entry (EAGLE) was developed by updating the drag acceleration periodically to generate corresponding angle of attack and bank angle [2]. Using the principle of linear quadratic regulator (LQR), longitudinal LQR-based tracking laws was presented in terms of linearized models [3, 4]. As a notable study in predictor corrector method, Fuhry has proposed a basic method of using error sensitivity coefficients to correct the predictions of trajectory [5], where only bank angle magnitude and sign reversals were controlled to meet terminal constraints. Zimmerman developed an improved predicted corrector approach where the bank angle profile was estimated with a profile follower using LQR [6].

Due to higher computational cost and convergence problem, predictor corrector methods are not applicable in practice in the near future. And as a profile tracking law, LQR method shows some merits over other methods, but produces tracking errors if aerodynamic uncertainties appear, which may lead to constraints violation or mismatched terminal conditions. Adding integral and differential terms can only make limited contribution to improve the performance due to model variation. Taking into account good generalization capability, fast learning ability and simple computation comparing with multilayer neural networks, fuzzy cerebellar model articulation controller (FCMAC) is introduced as auxiliary controller in longitudinal guidance. This integrated scheme corrects the command angle of attack and bank angle by tuning the weights online. In lateral profile, a crossrange parameter related corridor is derived according to nominal trajectory, and bank angle reversals are generated by dynamic adjust criterion to meet the state terminal requirements. Three dimension-of-freedom (3DOF) simulations are devoted to prove the effectiveness.

2 Nominal Profile of Lifting Reentry Vehicle

The model of lifting reentry vehicle is assumed to be an unpowered rigid point-mass in a stationary atmosphere with consideration of earth rotation. The dimensionless dynamics (1) consist of three kinematic equations and three force equations [7].

$$\begin{cases}
 \frac{dr}{d\tau} = V \sin \gamma \\
 \frac{d\theta}{d\tau} = \frac{V \cos \gamma \sin \psi}{R \cos \phi} \\
 \frac{d\phi}{d\tau} = \frac{V \cos \gamma \cos \psi}{R} \\
 \frac{dV}{d\tau} = \left(-D - \left(\frac{\sin \gamma}{R} \right) + \varphi_{v3} \right) \\
 \frac{d\gamma}{d\tau} = \frac{1}{V} \left[L \cos \sigma + \left(V^2 - \frac{1}{R} \right) \left(\frac{\cos \gamma}{R} \right) + \varphi_{\gamma3} + \varphi_{\gamma4} \right] \\
 \frac{d\psi}{d\tau} = \frac{1}{V} \left[\frac{L \sin \sigma}{\cos \gamma} + \frac{V^2 \cos \gamma \sin \psi \tan \phi}{R} - \varphi_{\psi3} + \varphi_{\psi4} \right]
 \end{cases} \tag{1}$$

where r is the radial distance of the vehicle from the center of the earth, normalized by the radius of the Earth $R_0=6378(\text{km})$. θ and ϕ are the longitude and latitude, respectively. V is the velocity of the vehicle, normalized by $V_c = \sqrt{g_0 R_0}$. γ is the flight path angle measured positive upward from the local horizontal plane. ψ is the clockwise angle from local north to velocity component normal to r . The bank angle σ is defined positive when turning right. $\tau = t / \sqrt{R_0 / g_0}$ as the differentiation variable represents the dimensionless time. The terms L and D are the dimensionless aerodynamic lift and drag accelerations. $L = \rho(V_c V)^2 S C_L / (2mg_0)$, $D = \rho(V_c V)^2 S C_D / (2mg_0)$, where S is the reference area of vehicle, m is the mass of vehicle, $\rho(r)$ is r dependent atmospheric density, $C_L(\alpha, Ma)$ and $C_D(\alpha, Ma)$ are the lift and drag coefficients, respectively. The functions φ_{v3} , $\varphi_{\gamma3}$, $\varphi_{\gamma4}$, $\varphi_{\psi3}$, $\varphi_{\psi4}$ are extra terms due to coriolis force and convected inertial force.

As shown in Fig.1, due to the configuration of lifting body, lifting reentry vehicle has large lift-to-drag ratio and shows gentler and longer-distance trajectory comparing with ballistic vehicle. So the key problem of nominal trajectory method is how to dissipate large orbital energy with demanded downrange and constraints. After giving a reentry height-velocity profile including overload N_{max} , heating rate Q_{smax} , dynamic pressure q_{max} constraints and quasi-equilibrium glide condition (QEGC), also with terminal radial distance r_f , range-to-go s_f and velocity V_f , and the heading error angle $\Delta\psi_f = \psi - \psi_{LOS}$, an optimal control problem with state variables, control inputs and terminal constraints is derived. While the angle of attack α is scheduled as a function of V and initial states are given, the proposed optimal control problem can be transferred into a constrained parameters optimization problem with discrete states and control inputs. The solution can be formulated by using conjugate gradient algorithm, sequence quadratic programming (SQP) algorithm or other methods. Fig.2 shows one reference height-velocity profile, where the shaded parts are upper and lower boundaries of the reentry corridor.

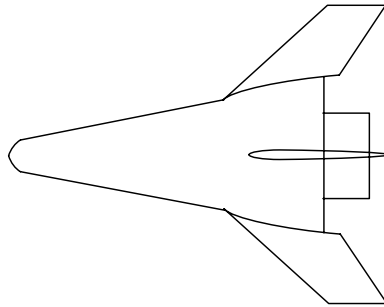


Fig. 1. The configuration of lifting reentry vehicle

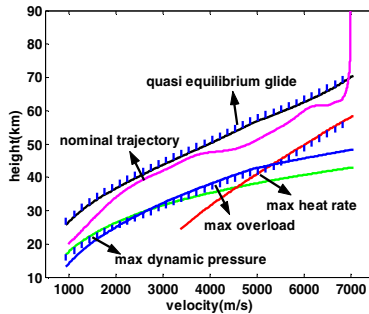


Fig. 2. Reference height profile and reentry corridor

3 Reentry Longitudinal Guidance Law

3.1 LQR Based Longitudinal Guidance

Because most of the critical trajectory parameters, such as downrange, energy, and path-constraint observance are concerned with longitudinal motion, an adaptive strategy is

called for tracking the nominal longitudinal profile. The drag acceleration tracking approach is not of high accuracy for the reason that relationship among the aerodynamic parameters of the lifting reentry vehicle is rather complicated. If uncertainties and aerodynamic dispersions exist, strict terminal constraints are hard to satisfy. In order to get better performance on longitudinal guidance, the small deviation linearized model to design linear controller is presented as follow:

$$\delta\dot{x} = A\delta x + B\delta u \tag{2}$$

where $\delta x = [r - r_{ref}, V - V_{ref}, \gamma - \gamma_{ref}, s - s_{ref}]^T$, s stands for the range-to-go with the definition $ds/d\tau = -V \cos \gamma \cos \Delta\psi / r$, and subscript ‘ref’ represents the nominal trajectory derived in section 2, $\delta u = [\delta\alpha, \delta\sigma]^T = [\alpha - \alpha_{ref}, \sigma - \sigma_{ref}]^T$. A and B are time-varying matrices obtained at given points by calculating the partial derivatives of the transformed differential equations along the nominal trajectory. It is known that additional robustness can be obtained by changing the independent variable from time to monotonic variable related to the vehicle states. So after dimensionless negative energy $e = 1/r - V^2/2$ is substituted into (2) as the independent monotonic variable, the initial and final energy are both fixed, and the infinite time optimal control problem becomes a two-point boundary value problem. The linearized model after transformation is

$$\delta x / \delta e = A' \delta x + B' \delta u \tag{3}$$

As a result of the relationship between the radial distance and velocity built through the independent variable e , V will be satisfied automatically by the constraints on radial distance r and can be omitted in the state vector x . So the optimal performance criterion is given as follow:

$$J = \int_{e_0}^{e_f} [\delta x^T(e) Q \delta x(e) + \delta u^T(e) R \delta u(e)] de / 2 \tag{4}$$

where $Q^{3 \times 3}$ and $R^{2 \times 2}$ are the weighting matrices used to tradeoff tracking accuracy versus control effort. The linear state feedback control law can be expressed as

$$\delta u = [\delta\alpha \ \delta\sigma]^T = -R^{-1} B'^T P \delta x(e) = -K' \delta x \tag{5}$$

where $K' \in R^{2 \times 3}$ is the gain matrix; the symmetric matrix P is the solution of Riccati equation $PA' + A'^T P - PB'R^{-1}B'^T P + Q = 0$.

Because the dynamics vary widely over the entry trajectory, linearized matrices A and B are different from each other at operating points along the nominal trajectory, so as the gain matrix K' . When the energy-scheduled strategy is applied, the state feedback gain coefficients in K' are got from the offline look-up table obtained in advance. The preliminary simulation results in [3] show that a set of appropriate gain matrices calculated along a particular nominal trajectory are insensitive to changes in different initial conditions and different nominal trajectories.

3.2 Description of FCMAC

Owing to the fact that LQR control law is derived through a set of linear models along one nominal trajectory, even if the above guidance law shows robust performance, it would still produce apparent tracking errors or terminal states deviation when

aerodynamic dispersions or uncertainties appear, which will make the value of heat rate, overload or dynamic pressure violate the pre-calculated corridor.

In this section, an integrated scheme comprised of LQR and FCMAC is proposed. FCMAC presented by Chiang and Lin [8] is a kind of adaptive neural network based on table look-up manner, which possesses nonlinear function approximation, fast-learning ability and good generalization capability. It uses Gaussian basis function to preserve the derivative information and improve the accuracy of the network output. The k^{th} receptive field function and the corresponding multidimensional vector can be expressed as:

$$\begin{cases} \Phi_k = \prod_{i=1}^n \phi_{A_{ik}}(x_i) = \exp\left[-\sum_{i=1}^n \frac{(x_i - m_{ik})^2}{\sigma_{ik}^2}\right] \\ \Phi = [\Phi_1, \Phi_2, \dots, \Phi_{n_R}]^T \end{cases} \quad (6)$$

where x_i is the normalized input, m_{ik} and σ_{ik} denote mean and variance of the membership function, n_R is the number of receptive fields. The output of FCMAC represents the algebraic sum of activated weights and receptive field value, which can be expressed in a vector form:

$$y = [y_1, y_2, \dots, y_m]^T = W^T \Phi \quad (7)$$

where matrix W represents the connecting weight values of the output associated with the receptive-fields. The network structure is shown as Fig. 3.

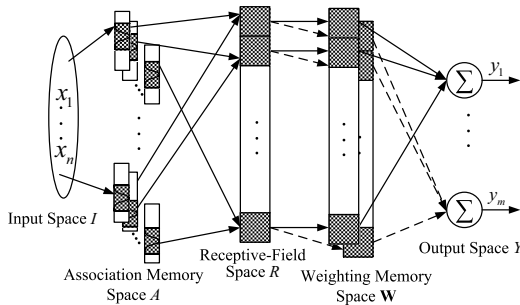


Fig. 3. General architecture of FCMAC

3.3 Improved Reentry Longitudinal Guidance Base on FCMAC

In this section, taking into account the mechanism of energy-scheduled longitudinal guidance law, FCMAC is devoted to be an auxiliary controller to enhance tracking performance of the nominal trajectory. In the integrated control structure, LQR controller is utilized to guarantee the stability of the tracking system. FCMAC serves as a compensator to make range-to-go and radial distance track the reference precisely and tunes the weights online in accordance with tracking errors.

In view of the great influence on flight height and flight path angle made by the bank angle σ and the variation of drag force impacted by the angle of attack α , we select range-to-go error e_1 , radial distance error e_2 and the corresponding derivatives

as components of two separate FCMAC networks' inputs. Furthermore, the two outputs would be added to the nominal control inputs σ_{ref} and α_{ref} , to eliminate nonlinearity and uncertainties contained in the linearized guidance model, which will help to improve the tracking effect of range-to-go and the radial distance. The control structure of the integrated guidance with LQR and FCMAC is shown in Fig. 4.

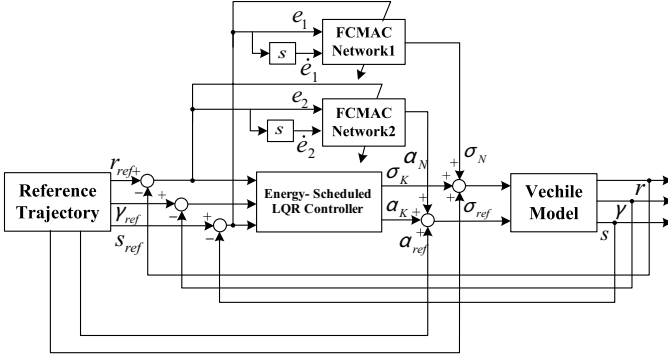


Fig. 4. Control structure of the integrated guidance with LQR and FCMAC

The control algorithm is given by the following equations:

$$\begin{cases} u_N(k) = W^T \Phi \\ u(k) = u_{ref}(k) + u_K(k) + u_N(k) \end{cases} \quad (8)$$

where $u_{ref}(k)$ is control input vector from the nominal trajectory, $u_N(k)=[\alpha_N(k) \ \sigma_N(k)]^T$ is the vector composed of two FCMAC output components, and $u_K(k)$ is the output vector produced by LQR controller.

Considering the compensating role FCMAC networks play in the integrated control, the traditional feed forward structure in FCMAC network is not concerned for its complexity and unavailability. Instead, the feedback scheme is presented in the closed-loop system, where dynamic errors are employed as input signals, and gradient descent method is adopted as weights adjusting regulation. According to the demand on minimum tracking error, the weight tuning law is given by:

$$\omega_j(k) = \omega_j(k-1) + \kappa(y(k) - y_d(k))\Phi_j \quad (9)$$

where κ is learning rate, j denotes the corresponding memory address activated by the k^{th} sample, Φ_j is the value of j^{th} receptive field.

4 Lateral Guidance Law

The aim of lateral guidance is to gain the sign of σ by tracking the longitudinal profile. The conventional bank-reversal criterion depends on whether the condition $|\Delta\psi| \leq \Delta\psi_{threshold}$ is observed. If not, the sign of the bank angle is set to the opposite of the current heading error, which would be maintained until the condition is violated again. It works well when the actual flight goes as planned in offline optimization. In

case the threshold value $\Delta\psi_{threshold}$ is given unreasonably, the sign of bank angle needs adaptive adjustments. Otherwise, frequent bank reversals or exceeded terminal heading errors out of the prespecified threshold may be obtained.

A crossrange parameter χ_e is defined as $\chi_e = \sin^{-1}[\sin(s)\sin\Delta\psi]$. This crossrange parameter represents magnitude of range-to-go, as well as large and fast variation of the heading error. Since the equation $ds/de = -\cos\gamma\cos\Delta\psi/rD$ exists, crossrange parameter χ_e is almost linear with respect to negative energy e . Once the sign of bank angle is reversed, the value of $d\Delta\psi/de$ and $d\chi_e/de$ will change accordingly. This characteristic can be conveniently applied in designing the bank-reversal criterion.

The bank-reversal times are pre-determined before optimization algorithms are devoted to generate nominal trajectory. In view of the principle that unnecessary bank reversals should be avoided, two bank reversals are chosen to ensure terminal constraints. $\Delta\psi_{ref}$ is assumed to be the heading error angle at the second bank reversal along the nominal trajectory and the lateral bound is defined as $\chi_{bound} = \sin^{-1}(\sin(s)\sin\Delta\psi_{ref})$. From [9], it is known that location of the first bank reversal is critical to lateral guidance law. Once the location takes place too early, an excessive number of σ will be obtained. But if the location is set to a much latter position, the terminal heading condition may not be met. Here, the first bank reversal is determined by:

$$|\chi_e| > c_1 |\chi_{bound}| \tag{10}$$

where $\eta = (L/D)_{est}/(L/D)_{ref}$, and c_1 is a scale coefficient linearly related to η . The less c_1 is, the earlier the location of the first bank reversal is. Since the crossrange parameter χ_e is approximately linear with respect to energy e , location of the second and latter reversals can be determined by the following equation:

$$|\chi_e| \geq |\chi'_e/\chi'_{bound}| |\chi_{bound}| \tag{11}$$

where $\chi'_e = d\chi_e/de$, $\chi'_{bound} = d\chi_{bound}/de$. Once the condition (11) is satisfied, the second reversal would take place instantly. If the linear relationship is not evident, third or latter bank reversals are generated near the terminal interface according to condition (11) to correct actual trajectory return to zero at e_f , which will guarantee χ_e back to permissible boundary. Fig.5 shows the cases of bank reversals.

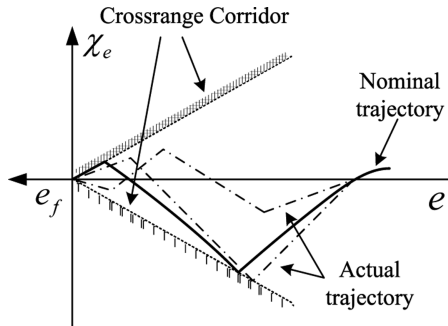


Fig. 5. Crossrange corridor used for bank angle reversals. The reference crossrange parameter variation is shown as a solid line. The actual crossranges are displayed in dashed-point line.

5 Simulation Results and Analysis

As to a certain type of lifting reentry vehicle, 3DOF numerical simulations have been carried out to validate the improved guidance law. LQR controller and integrated controller, which combine LQR method with FCMAC, are adopted respectively in tracking nominal trajectory. Dynamic adjustment method presented in section 4 is introduced as regulation to determine the bank reversals inside the crossrange corridor.

The initial states of the vehicle are $[h_0, \theta_0, \phi_0, v_0, \gamma_0, \psi_0] = [90\text{km}, 0^\circ, 0^\circ, 7000\text{m/s}, 0^\circ, 90^\circ]$, and the weighting matrices of LQR controller in the simulation is set to:

$$Q = \begin{bmatrix} 100 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 500 \end{bmatrix}, R = \begin{bmatrix} 50000 & 0 \\ 0 & 5000 \end{bmatrix}$$

Then the obtained gain matrices K' are used for interpolation by negative energy e . At different operating points, corresponding LQR gains are used. For the input space of FCMAC networks, the number of dimension $N = 20$. Due to the dimensionless inputs variation range, the initial values of m_{ij} for the receptive-field functions are given as equally division points from -0.08 to 0.08, and $\sigma_{ij} = 0.0084$ for all i and j . The initial weight matrix of the network is set to 0, learning rate of the network $\kappa = 0.1$ and simulation step is 0.1s. In lateral guidance law, $\Delta\psi_{ref} = 0.42(\text{rad})$, $c_1 = 0.2 + (\eta - \eta_{\min}) / (\eta_{\max} - \eta_{\min}) \times 0.8$, $\eta_{\min} = 0.6$, $\eta_{\max} = 1.4$. With a 10% dispersion in L/D , height tracking profiles and range-to-go tracking profiles with LQR controller and integrated controller are shown respectively in Fig. 6-a)~6-d). The histories of the bank angle and the angle of attack are plotted in Fig.7. Crossrange parameter χ_e and heading error $\Delta\psi$ variations are depicted in Fig. 8 and Fig. 9 separately. Statistics on terminal conditions are given by Table 1 for each case, where $\Delta h_f = h_f - h_{f_actual}$, $\Delta v_f = v_f - v_{f_actual}$, $\Delta s_f = s_f - s_{f_actual}$, $\Delta\psi_f = \psi_f - \psi_{f_actual}$. All these definitions are self-explained.

Table 1. Terminal conditions of the proposed guidance law in L/D deviation cases

L/D Dispersion	Control methods	Δh_f (km)	Δv_f (m/s)	Δs_f (km)	$\Delta\psi_f$ ($^\circ$)
10%	LQR control	-2.359	-90.29	11.23	3.631
	Integrated control	-0.358	-4.732	3.373	-0.263
-10%	LQR control	3.059	83.51	-15.43	1.347
	Integrated control	0.447	12.48	-1.475	0.225

Fig. 6-a)~6-d) reveal that, due to dispersion of L/D , there are evident errors between actual profile and command profile if LQR controller is applied, while both height tracking error and range-to-go tracking error are alleviated remarkably by integrated control with online learning of FCMAC network, just as the same outcome presented in Table 1. The view of terminal conditions summarizes that χ_e and $\Delta\psi$ both return to a small neighborhood of zero when 10% dispersion in L/D appears. It validates the effectivity of this dynamic adjustment criterion by determining two or more bank reversals autonomously. So the effectiveness and robustness of this proposed scheme for dispersion are verified.

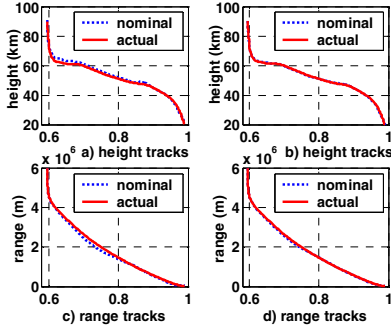


Fig. 6. Tracking profile in 10% *L/D* dispersion case. a), b) are height tracks by LQR method and integrated method respectively. c), d) are range-to go tracks by the above two methods accordingly.

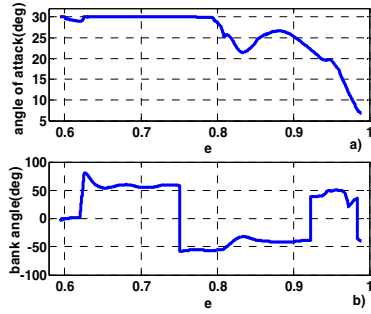


Fig. 7. a) angle of attack in 10% *L/D* dispersion case b) bank angle in 10% *L/D* dispersion case

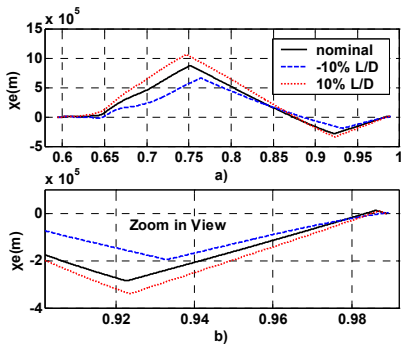


Fig. 8. a) Crossrange parameter along the nominal trajectory in 10% *L/D* deviation cases b) Zoom in view of crossrange χ_e

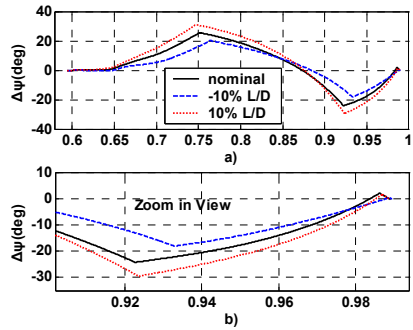


Fig. 9. a) Heading errors along the nominal trajectory in 10% *L/D* deviation cases b) Zoom in view of heading error

6 Conclusion

An integrated guidance law for lifting reentry vehicle is proposed in this paper. LQR method is developed to generate energy scheduled feedback gain matrices to form the main part of longitudinal guidance law, while FCMAC networks are introduced into the control loop as compensators to eliminate the errors made by aerodynamic uncertainties. In lateral guidance, the sign of bank angle is determined according to dynamic adjusting criterion inside a crossrange parameter corridor. 3DOF numerical simulations of a certain lifting reentry vehicle show satisfactory terminal and path conditions in presence of aerodynamic uncertainties. So reliability and validity of the presented method is demonstrated.

Acknowledgments. This work was supported in part by National Nature Science Foundation of China under Grant 60674105, Ph.D. Programs Foundation of Ministry of Education of China under Grant 20050487013, and the Nature Science Foundation of Hubei Province of China under Grant 2007ABA027.

References

1. Zhao, H.Y.: Reentry Dynamics and Guidance for Aircrafts. National University of Defense Technology Press, Changsha (1997)
2. Leavitt, J.A., Saraf, A., Chen, D.T.: Performance of Evolved Acceleration Guidance Logic for Entry (EAGLE). In: AIAA Guidance, Navigation and Control Conference and Exhibit, Monterey, California (2002)
3. Dukeman, G.A.: Profile-Following Entry Guidance Using Linear Quadratic Regulator Theory. In: AIAA Guidance, Navigation and Control Conference and Exhibit, Monterey, California (2002)
4. Ning, G.-d., Zhang, S.-g., Fang, Z.-p.: Integrated Entry Guidance for Reusable Launch Vehicle. *Chinese Journal of Aeronautics* 20, 1–8 (2007)
5. Fuhry, D.P.: Adaptive Atmospheric Reentry Guidance for the KistlerK-1 Orbital Vehicle. In: AIAA Guidance, Navigation, and Control Conference and Exhibit, Collection of Technical Papers, Portland, OR, vol. 2 (1999)
6. Zimmerman, C., Dukeman, G., Hanson, J.: An Automated method to compute orbital reentry trajectories with heating constraints. *Journal of Guidance, Control, and Dynamics* 26, 523–529 (2003)
7. Shen, Z., Lu, P.: On-Board Generation of Three-Dimensional Constrained Entry Trajectories. *Journal of Guidance, Control, and Dynamics* 26, 111–121 (2003)
8. Chiang, C.T., Lin, C.S.: CMAC with general basis functions. *Neural Networks* 9, 1199–1211 (1996)
9. Shen, Z., Lu, P.: Lateral Entry Guidance Logic. *Journal of Guidance, Control, and Dynamics* 27, 949–959 (2004)

Hybrid Filter Based Simultaneous Localization and Mapping for a Mobile Robot

Kyung-Sik Choi, Bong-Keun Song, and Suk-Gyu Lee

Department of Electrical Engineering, Yeungnam University,
214-1 Daedong Gyongsan Gyongbuk Republic of Korea
robotics@ynu.ac.kr, 01071630579@nate.com, sglee@ynu.ac.kr

Abstract. We propose a hybrid filter based SLAM (Simultaneous Localization and Mapping) for a mobile robot to compensate for the EKF (Extended Kalman Filter) based SLAM error inherently caused by the linearization process. A mobile robot autonomously explores the environment by interpreting the scene, building an appropriate map, and localizing itself relative to this map. A probabilistic approach has dominated the solution to the SLAM problem. This solution is a fundamental requirement for robot navigation. The EKF algorithm with a RBF (Radial Basis Function) has some advantages in handling a robotic system having nonlinear dynamics because of the learning property of neural networks. We modified an already developed Matlab simulation source for the hybrid filter-SLAM for simulation and comparison. The simulation results showed the effectiveness of the proposed algorithms as compared with an EKF-based SLAM.

Keywords: SLAM, Hybrid filter, Neural networks, EKF, Mobile robot, RBF algorithm.

1 Introduction

Research efforts on mobile robotics have mainly focused on topics such as autonomous navigation, path planning, map-building, etc. [1]. Currently SLAM is one of the most widely researched major subfields of mobile robotics. It is a relatively new sub-field of robotics. In order to solve SLAM problems, statistical approaches such as Bayesian Filters have received widespread acceptance. Some of the most popular approaches for SLAM include using a Kalman filter (KF), an extended Kalman filter (EKF), and a particle filter [2]-[4]. The earliest SLAM algorithm was based on the extended Kalman filter. This algorithm has been applied with some success in practice. As any EKF algorithm, EKF-SLAM makes a Gaussian noise assumption for robot motion and perception. In addition, the amount of uncertainty in the posterior of the EKF SLAM algorithm must be relatively small; otherwise, the linearization in the EKFs tends to introduce intolerable errors. Differently from the EKF, the main objective of particle filtering is to track a variable of interest as it evolves over time, typically with a non-Gaussian and potentially a multi-modal probability density function (PDF). The introduction of particle filters gave researchers the power and flexibility to handle nonlinearity and non-Gaussian distributions routinely. The basis of the

method is to construct a sample-based representation of the entire PDF, a main difference when comparing with an EKF using parameterization. A neural network, adaptive to changes of environmental information flowing through the network during the process, can be combined with a EKF to compensate for some of the disadvantages of a EKF-SLAM which represents the state uncertainty by an approximate mean and variance and has biased systematic errors even after appropriate compensation in real situations[4]-[9].

Choi, *et al* [4] approached the SLAM problem with neural networks aided with an extended Kalman filter (NNEKF) by comparing the EKF-SLAM with a NNEKF (neural network-aided extended Kalman Filter). Research has shown the NNEKF-SLAM with better performance than the EKF-SLAM, when they used multi layer perceptrons (MLP). Stubberud *et al* [10] developed an adaptive EKF using artificial neural networks. Previous work on the EKF-SLAM shows that an eventual inconsistency of the algorithm is inevitable for large-scale maps.

In this paper, we discuss our use of MLP (multilayered perceptron) and a radial basis function (RBF) algorithm to handle nonlinear properties of a mobile robot. We propose a hybrid filter-SLAM (MLP-SLAM and RBF-SLAM) to reduce the estimation error comparing with an EKF-SLAM, often considered a standard SLAM approach.

2 Related Algorithms for SLAM

2.1 Neural Networks

In this paper, we describe our use of two types of neural networks: the Multi Layer Perceptrons Algorithm (MLP) and the Radial Basis Function Algorithm (RBF). The MLP, having hidden layers with one or more input and output nodes, is a typical feed-forward neural network model used as a universal approximator [10]-[12]. Output signals are generated through the homogeneously nonlinear function after summing signal values for each of the input nodes [8]. In this process, signals are multiplied by weights and added by bias values. The RBF network uses radial basis functions as activation functions for function approximation, control, etc. RBF networks typically have three layers: an input layer, a hidden layer with a nonlinear RBF activation function, and a linear output layer. Network training is divided into two stages: First, the weights from the input to the hidden layer are determined; then, the weights from the hidden to the output layer are determined. The results can be used to simulate the nonlinear relationship between the sensors' measurements with errors and the ideal output values by using the least squares method [5][9].

2.2 EKF-SLAM

A solution to the SLAM problem using an EKF, with many interesting theoretical advantages, is probably described the most in research literature. This is despite the recently reported inconsistency of its estimation because it is a heuristic for the nonlinear filtering problem. Associated with the EKF is the Gaussian noise assumption. This assumption significantly impairs the EKF SLAM's ability to deal with uncertainty. With a greater amount of uncertainty in the posterior, the linearization in

the EKF fails. An EKF based on a Bayes filter has two steps, prediction and update, for SLAM using the measured sensor data of a mobile robot [10][13][14].

3 A Hybrid Filter SLAM Algorithm

We propose a hybrid filter-SLAM with an EKF augmented by an artificial neural network (NN) acting as an observer to learn the system uncertainty on-line. We developed an adaptive state estimation technique using an EKF and a NN. In this research, the mobile robot with encoder values (u_t, ω_t) learns values (x'_t, y'_t, θ'_t) that are driven from environmental information values (x_t, y_t, θ_t) using MLP and RBF algorithms as shown Fig. 1.

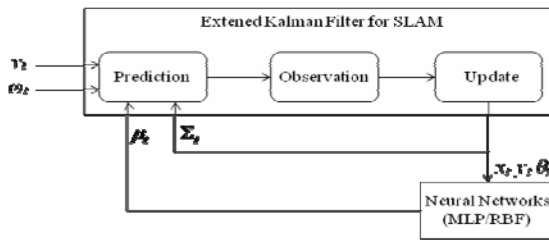


Fig. 1. The architecture of a Hybrid filter SLAM

3.1 A Motion Model for the SLAM

We describe the hybrid filter-SLAM algorithm using a robot’s pose: $(x^R \ y^R \ \theta^R)$ and its features such as location of landmarks: $(x_{Lx} \ x_{Ly})$. The estimated error covariance is defined as in Eq. (1), where the diagonal sub-matrices are a covariance of the vehicle and its features. The off-diagonal-matrices are their correlation.

$$\Sigma = \begin{pmatrix} \Sigma_{vv} & \Sigma_{vL1} & \dots & \Sigma_{vLn} \\ \Sigma_{L1v} & \Sigma_{L1L1} & \dots & \Sigma_{L1Ln} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{Lnv} & \Sigma_{LnL1} & \dots & \Sigma_{LnLn} \end{pmatrix} \tag{1}$$

We can present the state transition probability of a hybrid filter-SLAM in the linearity assumption where g represents nonlinear functions, ε is process noise, and u_t is the control command with velocity v_t , and an angular velocity ω_t for the mobile robot.

$$x_t = g(x_{t-1}, u_t) + \varepsilon_t \tag{2}$$

$$u_t = \begin{pmatrix} v_t \\ \omega_t \end{pmatrix} \tag{3}$$

For the Taylor expansion of function g , we use its partial derivative as shown in Eq.(4).

$$g'(x_{t-1}, u_t) = \frac{\partial g(x_{t-1}, u_t)}{\partial x_{t-1}} \tag{4}$$

g is approximated at μ_{t-1} and u_t . The linear extrapolation is achieved by using the gradient of g at μ_{t-1} and u_t .

$$\begin{aligned} g(x_{t-1}, u_t) &\approx g(\mu_{t-1}, u_t) + g'(\mu_{t-1}, u_t)(x_{t-1} - \mu_{t-1}) \\ &= g(\mu_{t-1}, u_t) + G_t(x_{t-1} - \mu_{t-1}) \end{aligned} \tag{5}$$

G_t , a Jacobian, is a matrix with dimension $n \times n$, where n denotes the dimension of the state. It has a different value at each μ_{t-1} and u_t .

$$G_t = \frac{\partial g(\mu_{t-1}, u_t)}{\partial x_{t-1}} \tag{6}$$

3.2 The Measurement Step of Probability

The measurement probability consists of the nonlinear measurement function h and the observation noise δ :

$$z_t = h(x_t) + \delta_t \tag{7}$$

Since the measurement function h is an expansion of g , the Taylor expansion is developed around $\bar{\mu}_t$.

$$h'(x_t) = \frac{\partial h(x_t)}{\partial x_t} \tag{8}$$

$$h(x_t) \approx h(\bar{\mu}_t) + h'(\bar{\mu}_t)(x_t - \bar{\mu}_t) = h(\bar{\mu}_t) + H_t(\bar{\mu}_t)(x_t - \bar{\mu}_t) \tag{9}$$

The Jacobian H_t of the measurement function h is calculated at the predicted mean $\bar{\mu}_t$.

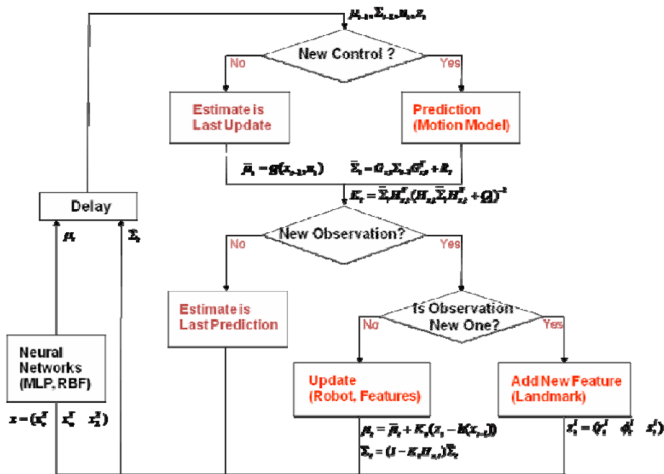


Fig. 2. A flowchart of the hybrid filter SLAM algorithm

3.3 The Predict Step of the Hybrid Filter SLAM

We applied the NN algorithm to the observation step of the EKF-SLAM to lessen the error of the mobile robot's pose. The prior mean $\bar{\mu}_i$ and covariance $\bar{\Sigma}_i$ have the form of

$$\bar{\mu}_i = \mathbf{g}(x_{i-1}, u_i) \quad (10)$$

$$\bar{\Sigma}_i = G_{x,i} \Sigma_{i-1} G_{x,i}^T + R_i \quad (11)$$

The motion model requires motion noise to be mapped into state space. The Jacobian needed for the approximation, denoted as V_i , is the derivative of the motion function g , with respect to the motion parameters, evaluated at μ_{i-1} and u_i .

$$V_i = \frac{\partial \mathbf{g}(\mu_{i-1}, u_i)}{\partial u_i} \quad (12)$$

To derive the covariance of the additional motion noise, $N(0, R_i)$, we need to know the covariance matrix M_i of the noise in control space. In addition, we can represent the motion noise R_i by multiplying the Jacobian V_i and the covariance matrix M_i of the noise in control space.

$$M_i = \begin{pmatrix} \alpha_1 v_i^2 + \alpha_2 \omega_i^2 & 0 \\ 0 & \alpha_3 v_i^2 + \alpha_4 \omega_i^2 \end{pmatrix} \quad (13)$$

$$R_i = V_i M_i V_i^T \quad (14)$$

3.4 The Observation Step of the Hybrid Filter SLAM

To derive the Kalman gain K_i , we confirm measurement noise covariances and the measurement model z_i^i for feature-based maps. Let $j = c_i^i$ be the identity of the landmark that corresponds to the i -th component in the measurement vector. c_i^i is a set of correspondence variables that is the true identity of an observed feature.

$$z_i^i = \begin{pmatrix} r_i^i \\ \phi_i^i \\ s_i^i \end{pmatrix} = h(x_i, j, m) + N(0, Q_i) \quad (15)$$

$$H_i^i = \frac{\partial h(\bar{\mu}_i, j, m)}{\partial x_i} \quad (16)$$

Using the covariance Q_i , a diagonal matrix with elements of z_i^i , the Kalman gain has the form of Eq. (17).

$$K_i = \bar{\Sigma}_i H_{x,i}^T (H_{x,i} \bar{\Sigma}_i H_{x,i}^T + Q_i)^{-1} \quad (17)$$

3.5 The Update Step of the Hybrid Filter SLAM

In the update step, the mean μ_i and the covariance Σ_i of the measurement are updated in terms of the prior mean $\bar{\mu}_i$ and the prior covariance $\bar{\Sigma}_i$.

$$\mu_t = \bar{\mu}_t + K_t(z_t - h(x_{t-1})) \tag{18}$$

$$\Sigma_t = (I - K_t H_{x,t}) \bar{\Sigma}_t \tag{19}$$

To apply a NN, we divide the mean values of the measurement into each component, x^R , y^R , θ^R , and the input NN (MLP and RBF) algorithm. In a MLP with two hidden layers, we assume it to have no bias. Eq. (22) describes a component such as $x_t^0 = x_t^R$, $x_t^1 = y_t^R$ and $x_t^2 = \theta_t^R$ where j, k, l are the numbers of a layer's node, respectively.

$$y_t^j = \xi \left(\sum_{l=0}^{p-1} w_t^{kl} \phi_t^{k^l} \right) = \xi \left(\sum_{l=0}^{p-1} w_t^{kl} \xi \left(\sum_{j=0}^{l-1} w_t^{jk} \phi_t^j \right) \right) = \xi \left(\sum_{l=0}^{p-1} w_t^{kl} \xi \left(\sum_{j=0}^{l-1} w_t^{jk} \xi \left(\sum_{i=0}^{M-1} w_t^{ij} x_t^i \right) \right) \right) \tag{20}$$

(0 ≤ j ≤ 6, 0 ≤ k ≤ 6, 0 ≤ l ≤ 2)

In the simulation using the second algorithm based on a RBF with 25 neurons, x_t is a n -dimensional input vector and c_i is the center of the i -th basis function with the same dimension of the input vector x_t . In addition, we have some kinds of elements such as, σ^i : the width of the basis function, N : the number of hidden layers, $\|x_t^i - c^i\|$: the norm of $x_t^i - c^i$ representing the distance between x and c^i , and $\phi^i(x)$: the response of the i -th basis function of the input vector with a maximum value at c^i .

$$y_t^i = \xi \left(\sum_{i=0}^{N-1} \phi^i(x_t^i) \right) = \xi \left(\sum_{i=0}^{N-1} \exp \left(-\frac{\|x_t^i - c^i\|^2}{2(\sigma^i)^2} \right) \right) \quad (0 \leq N \leq 25) \tag{21}$$

Finally, we can substitute the derived result y_t^i for the measurement μ_t . The above 5 steps are repeated until the end of the mobile robot's exploration.

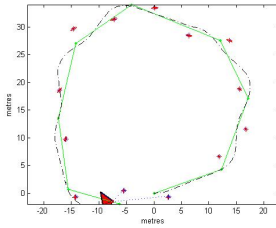
$$\mu_t = y_t^i \quad (x^R = y_t^0, y^R = y_t^1, \theta^R = y_t^2) \tag{22}$$

4 The Simulation Results and Discussion

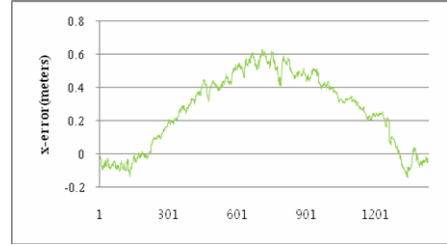
The simulation results, obtained by modifying the Matlab code developed by Bailey *et al* [14], show the efficiency of the proposed algorithm over the EKF-SLAM. The simulation was performed with constraints on velocity, steering angle, system noise, observation noise, etc., for a robot with a 2m-diameter wheel and 3m/sec maximum speed. The maximum steering angle and speed were 25° and 15°/sec respectively. The control input noise was assumed to be zero mean Gaussian with σ_v (=0.4 m/s) and σ_ϕ (=3°). The navigating environment was about 60 m × 40 m in a rectangular form. For observation, we used 13 features around waypoints such as 9 posed circles. In the observation step, we used a range-bearing sensor model and an observation model to measure the feature position and robot pose which had a noise level of 0.1 m in range and 1° in bearing. The sensor range was restricted to 20 m. This range was good enough to detect all features in front of the mobile robot.

4.1 A Simulation of an EKF-SLAM

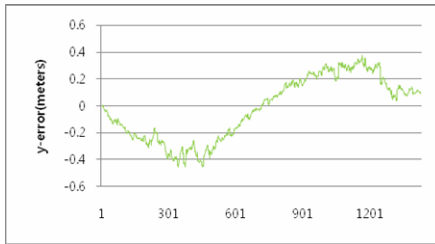
The amount of uncertainty in the posterior must be relatively small; otherwise, the linearization in the EKFs tends to introduce intolerable errors. Fig. 3(a) shows both the desired and EKF-based navigation trajectories of the mobile robot. The solid and dotted lines depict the reference and actual trajectories, respectively. The ellipse describes the estimated covariance of the mobile robot and features by the EKF. The plus (+) features are true and estimated features.



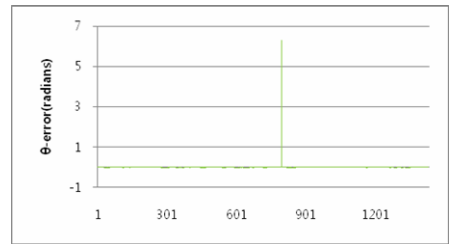
(a) The navigation error using the EKF-SLAM



(b) The x-axis error using the EKF-SLAM



(c) The y-axis error using the EKF-SLAM



(d) The θ error using the EKF-SLAM

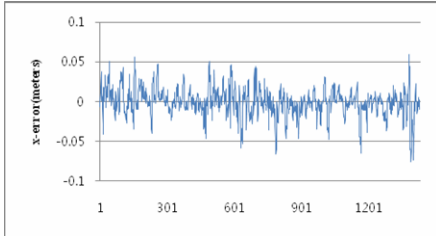
Fig. 3. Trajectory errors of the EKF-SLAM

The trajectory error of the EKF based simulation results mainly from the system errors and observation errors during the navigation of 1,433 time steps.

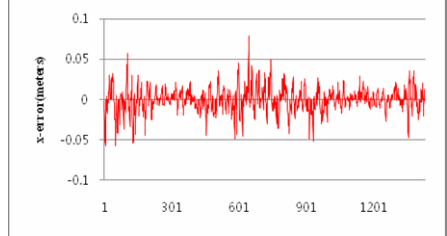
4.2 Simulation of the Hybrid Filter-SLAM

To enhance trajectory accuracy, we adopted a MLP with 3 inputs and outputs' nodes and 2 hidden layers with 7 nodes as a prior stage of the EKF. Each hidden layer has a sigmoid function. Fig. 4 shows navigation errors of the parameters of the hybrid filter-SLAM (MLP-EKF and RBF-SLAM). The MLP-EKF and RBF-EKF based simulation results show very similar performance in the x-axis and y-axis behavior. The latter results in better performance than the former in θ as shown in Fig. 4. In the RBF-EKF simulation, we used the same mean squared error of 0.001 as in the MLP-EKF.

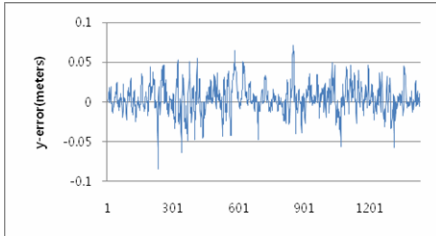
In general, both the MLP-EKF and RBF-EKF based SLAM reduce estimation errors as compared with the EKF SLAM. However, the training time in a neural network may result in some problems practically in real time operation. The training time in the RBF-SLAM is shorter than that in the MLP-SLAM and similar to the EKF-SLAM case.



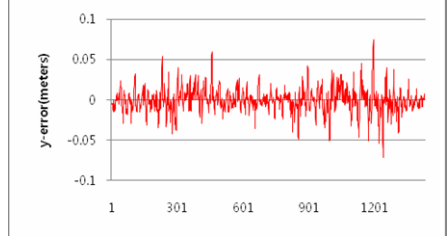
(a) The x-axis error using the MLP-EKF



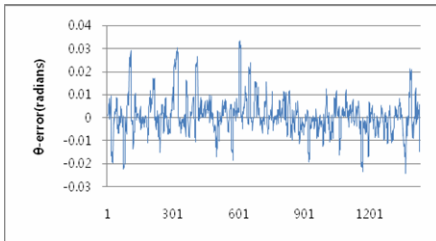
(b) The x-axis error using the RBF-EKF



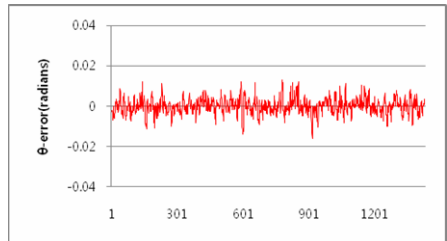
(c) The y-axis error using the MLP-EKF



(d) The y-axis error using the RBF-EKF



(e) The θ error using the MLP-EKF



(f) The θ error using the RBF-EKF

Fig. 4. Navigation errors in the parameters of the hybrid filter-SLAM (MLP-EKF and RBF-SLAM)

Fig. 5 shows simulation errors using the EKF, the MLP-EKF, and the RBF-EKF SLAMs. In Fig. 5(c), the RBF-EKF SLAM results in the smallest error in the heading error θ among the three methods.

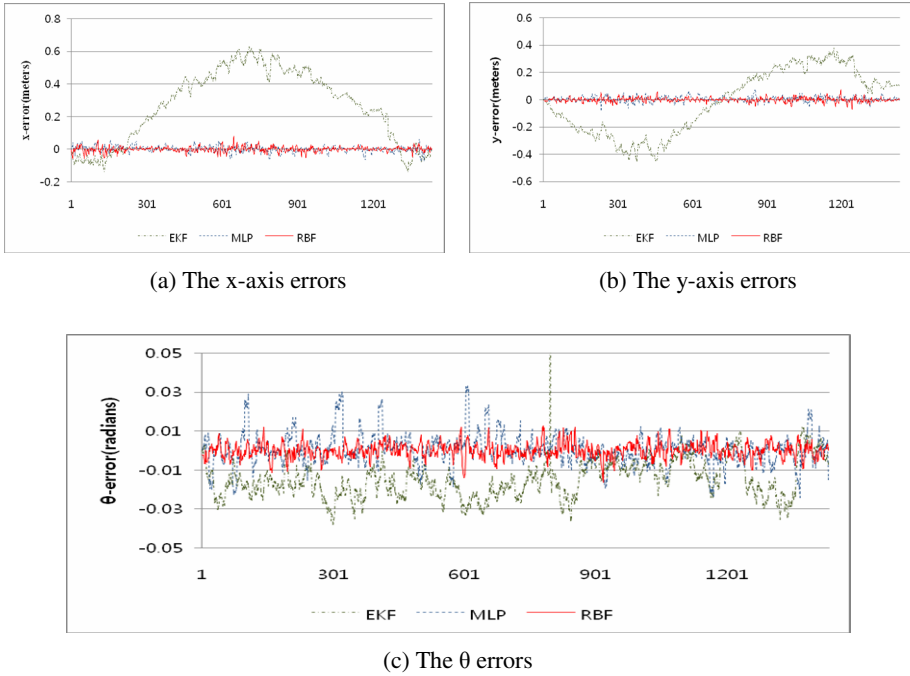


Fig. 5. Simulation errors : The EKF and hybrid filters

5 Conclusions

The SLAM problem is of fundamental importance in the quest for autonomous mobile robots since the robot keeps track of its location by maintaining a map of the physical environment and an estimate of its position on that map. In this paper, we propose hybrid filter-SLAMs, such as the MLP-SLAM and the RBF-SLAM of a mobile robot, to make up for the EKF-SLAM error inherently caused by its linearization process and noise assumption.

The proposed algorithm consists of two steps: the Neural Network and the extended Kalman Filter algorithm. The simulation results show the efficiency of the proposed algorithm as compared with the EKF-SLAM in terms of parameters such as x , y , and θ .

According to the simulation results, the RBF-SLAM shows the best performance in terms of trajectory error. For the real time realization of the proposed system, research on various training conditions and structures of neural networks, etc., is underway in our laboratory.

References

1. Kim, J.M., Kim, Y.T., Kim, S.S.: An Accurate Localization for Mobile Robot Using Extended Kalman Filter and Sensor Fusion. In: IEEE International Joint Conference on Neural Networks, pp. 2928–2933 (2008)

2. Panzieri, S., Pascucci, R., Setola, R.: Multirobot Localization Using Interlaced Extended Kalman Filter. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2816–2821 (2006)
3. Lee, S.J., Lim, J.H., Cho, D.W.: EKF Localization and Mapping by Using Consistent Sonar Feature with Given Minimum Features. In: SICE-ICASE International Joint Conference, pp. 2606–2611 (2006)
4. Harb, M., Abielmona, R., Naji, K., Petriul, E.: Neural Networks for Environmental Recognition and Navigation of a Mobile Robot. In: IEEE International Instrumentation and Measurement Technology Conference, pp. 1123–1128 (2008)
5. Zu, L., Wang, H.K., Yue, F.: Artificial Neural Networks for Mobile Robot Acquiring Heading Angle. In: Proceedings of the Third International Conference on Machine Learning and Cybernetics, pp. 26–29 (2004)
6. Bugeja, M.K., Fabri, S.G.: Multilayer Perceptron Adaptive Dynamic Control for Trajectory Tracking of Mobile Robots. In: IEEE Industrial Electronics Annual Conference, pp. 3798–3803 (2006)
7. Choi, M.Y., Sakthivel, R., Chung, W.K.: Neural Network-Aided Extended Kalman Filter for SLAM Problem. In: IEEE International Conference on Robotics and Automation, pp. 1686–1690 (2007)
8. Jang, P.S.: Neural Network Based Position Tracking Control of Mobile Robot. Chungnam National University, M.S These, pp. 13–37 (2003)
9. Oh, C.M.: Control of Mobile Robots Using RBF Network. Korea Advanced Institute of Science and Technology, M.S These, pp. 4–19 (2003)
10. Stubberud, S.C., Lobbia, R.N., Owen, M.: An Adaptive Extended Kalman Filter Using Artificial Neural Networks. In: Proceedings of the 34th Conference on Decision & Control, pp. 1852–1856 (1995)
11. Mehra, P., Wah, B.W.: Artificial Neural Networks: Concepts and Theory, pp. 13–31. IEEE Computer Society Press, Los Alamitos (1992)
12. Iiguni, Y., Sakai, H., Tokumaru, H.: A Real-Time Learning Algorithm for a Multilayered Neural Network Based on the Extended Kalman Filter. IEEE Transactions on Signal Processing 40(4), 959–966 (1992)
13. Thrun, S., Burgard, W., Fox, D.: Probabilistic Robotics, pp. 309–334. The MIT Press, London (2005)
14. Bailey, T., Nieto, J., Guivant, J., Stevens, M., Nebot, E.: Consistency of the EKF-SLAM Algorithm. In: IEEE International Conference on Intelligent Robotics and Systems, pp. 3562–3568 (2006)

Using Toe-Off Impulse to Control Chaos in the Simplest Walking Model via Artificial Neural Network

Saeed Jamali¹, Karim Faez², Sajjad Taghvaei³, and Mostafa Ozlati Moghadam⁴

¹ Department of Computer Engineering, Islamic Azad University Branch of Ghazvin, Ghazvin, Iran

s.jamali@gazviniau.ac.ir

² School of Electrical Engineering, Amirkabir University of Technology, Tehran, Iran

kfaez@aut.ac.ir

³ Department of Mechanical Engineering, Amirkabir University, Tehran, Iran

sj.taghvaei@aut.ac.ir

⁴ School of Engineering and Computer Science, Australian National University, Canberra, Australia

moghadam@ieee.org

Abstract. Controlling chaos in a passive biped robot with an artificial neural network is investigated in this paper. The dynamical model is based on the compass-like biped robot proposed by Garcia et al. (1998) with a point-mass at the hip and infinitesimal point-masses at the feet ignoring the scuffing situation. The governing dynamics and chaotic behavior of the system is explored and the bifurcation diagram is drawn with respect to the ramp slope. Controlling chaos is based on stabilizing the unstable periodic orbits in the chaotic attractor. The UPOs are detected using an iterated algorithm. The artificial neural network is constructed using the information of seven previous steps and the control parameters in each one. The network is trained to find the appropriate control parameter in order to put the next step on the unstable periodic orbit. The control parameter is the toe-off impulse at the heel strike.

Keywords: Artificial neural network, Chaos control, Passive biped, Simplest Walking model.

1 Introduction

Human locomotion is one of the most complicated dynamical motions in nature 13. That is why the phenomenon is not yet well understood despite being well investigated by the researchers. Investigating bipedal walking robots is shown to be an appropriate means to better understanding of anthropomorphic locomotion. Although human motion is controlled by the neuro-muscular system, except for part of a stride 14, McGeer (1990) showed that some legged mechanisms can exhibit stable human-like walking on a range of shallow ramps without actuation 1.

The compass-like biped robot investigated by Garcia et al. 1998 is assumed to be the simplest model capable of mimicking bipedal gait. The dynamics of the model has been studied widely by several researchers 15, 17, 16, 5, 18. Moreover several control

strategies have been applied to get a desired motion from the bipedal walkers. The control parameter is usually the torque at the joints 7 and is identified by different control algorithms and techniques. Linear control, based on linearization of the equations of motion around the vertical stance 8, variable structure control 9, optimal control 9, and shaping discrete event dynamics 11 are the main techniques that has been applied to biped robots by investigators.

The basic idea in controlling biped robots is choosing a proper control input in order to get a desired fashion. The desired behavior of a biped robot can be specified in numerous ways and still is an open question 19. Nevertheless one can assume a fundamental characteristic of a desired motion is that merely the biped doesn't fall.

Another common desired feature for bipedal walking with chaotic behavior is a stable periodic motion. The detection of limit cycles for the simplest walking model under various mechanical and environmental circumstances has been studied by several researchers 15, 5, 6.

Although there is an acceptable amount of research on controlling bipedal walkers with chaotic behavior, there is still few works on applying chaos control algorithms on the system. Chaos control algorithms have several major advantages due to dynamical properties of the system in the chaotic attractor. A chaotic system is described to be both flexible and stable 12, 22.

In 23 the authors have applied OGY algorithm 22 to a compass-like biped with masses on knees assuming the hip actuation torque as the controlling signal. The method was proposed by Ott, Grebogi and York in 1998 based on the linearization of the Poincare' map near the fixed point in the chaotic attractor. Kurz et al. have proposed a biologically inspired ANN algorithm to rapidly drive the trajectories to a stable periodic orbit 24. Actually the neural network proposed in 24 gives the proper spring stiffness for a single ramp incline and is said to be applied to the system after several steps. The control parameter is not adjustable enough and setting up a spring in the hip joint in the middle of walking seems not to be applicable. Moreover adding up the spring from the beginning actually changes the dynamics of the system and the chaotic characteristics of the model that have already been resulted through numerical manipulations.

In this investigation we have extended the Kurz's algorithm to control a passive biped robot using toe-off impulses as the control parameter. Choosing the toe-off impulse as the control actuator is thought to be more applicable and more similar to what occurs in human locomotion. The neural network is designed so that the controller can be turned on at any time during the locomotion. Moreover the network is trained to be applicable for any ramp slope.

The dynamic model of the system is first analyzed in section II. In section III, the bifurcation diagram is drawn and unstable periodic orbits are detected using an iterative algorithm proposed in 15. The artificial neural network is discussed in section IV and the simulation results are proposed in section V. Conclusions and suggestions for future work are the last section of this article.

2 Dynamic Model

The simplest walking model shown in Fig.1 is a biped robot with two rigid legs joint by a frictionless hinge at the hip. There is a point mass at the hip, M and two point

masses at the feet tips, m . The feet masses are much smaller than the hip so that the hip motion is not affected by the swinging leg motion. During each step the stance leg acts as an inverted pendulum while the other leg oscillates until the heel contact. At the moment the system has a plastic impact without any slip or bounce and passes a double-support phase instantaneously 15.

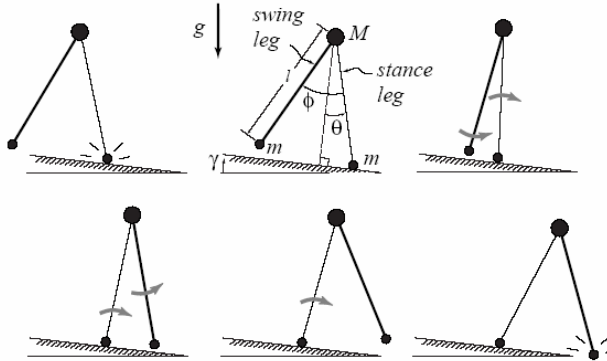


Fig. 1. The simplest walking model with a hip mass and two point masses at the feet tip 15. After the heel-strike the stance leg and the sing leg switch.

The legs are assumed to be able to pass the foot-scuff situation and the swing leg is able to momentarily pass through the ramp surface when the stance leg is nearly vertical. These assumptions are made to avoid scuffing problems for straight legged walkers.

The degrees of freedom are chosen to be θ and ϕ that shows the stance leg angle with the vertical direction to the ramp surface and the angle between stance and swing leg. The governing equation in the single-support phase is derived by Lagrange’s relations.

Substituting the Lagrangian of the system in the Lagrange’s relations, applying the simplifying assumption ($\frac{m}{M} \ll 1$) and rescaling time by $\sqrt{\frac{l}{\theta}}$, the governing equations become:

$$\begin{aligned} \ddot{\theta} - \sin(\theta(t)) - \gamma &= 0 \\ \ddot{\theta} - \ddot{\phi} + \dot{\theta}^2 \sin \phi(t) - \cos(\theta(t) - \gamma) \sin \phi(t) &= 0 \end{aligned} \tag{1}$$

During the single-support phase the above equation governs the motion of the biped. When the swing leg contacts to the surface satisfying the geometric condition

$$\phi - 2\theta = 0 \tag{2}$$

The collision occurs and solving differential equations (2) should be stopped. The conservation of angular momentum should be satisfied for the whole system about the swing foot tip and for the new stance leg around the hip. Considering the toe-off impulse denoted by P the transition rule 15 for heel strike moment is as follows.

$$\begin{aligned}
 \begin{bmatrix} \theta \\ \dot{\theta} \\ \phi \\ \dot{\phi} \end{bmatrix}^+ &= \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & \cos 2\theta & 0 & 0 \\ -2 & 0 & 0 & 0 \\ 0 & \cos 2\theta & (1-\cos 2\theta) & 0 \end{bmatrix} \begin{bmatrix} \theta \\ \dot{\theta} \\ \phi \\ \dot{\phi} \end{bmatrix}^- \\
 &+ \begin{bmatrix} 0 \\ \sin 2\theta \\ 0 \\ (1-\cos 2\theta) \sin 2\theta \end{bmatrix} P
 \end{aligned} \tag{3}$$

The states of the system just after and just before the impact are identified by ‘+’ and ‘-’ subscripts. The set of differential algebraic equations (1)-(3) are the governing equations of the simplest walking model.

3 Model Analyses

3.1 Poincare’ Section and Bifurcation Diagram

In order to analyze the dynamics of the model in a simpler way the state of the systems can be investigated stroboscopically by using a Poincare’ section. The Poincare’ section reduces one dimension of the system 12.

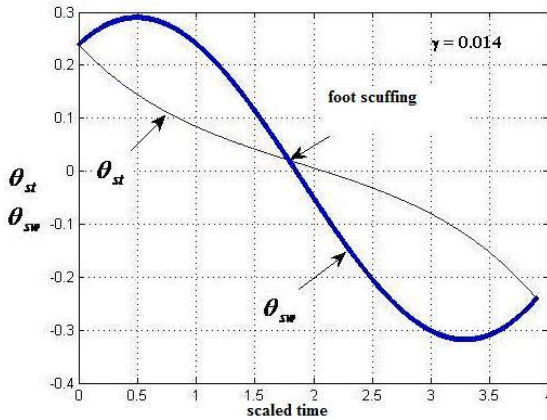


Fig. 2. The swing and stance leg angle variations for a typical step and the scuffing situation

The Poincare’ section chosen for this model is the one just after the heel strike situation. Since there is especial relations between the angular velocities at the moment (Eq. (4)) the Poincare’ map of the system would be 2D.

Analyzing the Poincare’ section data of the system shows that the simplest walking model has different stable walking patterns in different slope ranges. The system has a period doubling route to chaos if the slope changes continuously 15. The bifurcation

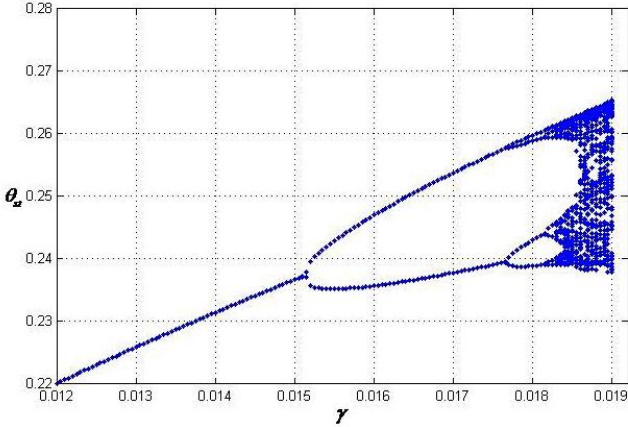


Fig. 3. Bifurcation diagram of the system with respect to the ramp slope [19]

diagram of the system plotted in figure (3) gives a good insight to the biped behavior and walking patterns.

For ($\gamma < 0.015$) there is a stable limit cycle which shows a periodic walking in which after a transient phase the steps are repeated. At $\gamma = 0.015$ the bifurcation occurs and the biped starts to show a limping motion. The period doubling continues until the system becomes chaotic. After there is no stable walking pattern and the biped falls with any initial condition.

3.2 Unstable Periodic Orbits

The unstable periodic orbits are said to be the skeleton of the chaotic attractor. The trajectories are attracted by and repelled from them in each cycle but never converge to them 26. It can be difficult to detect these orbits from an observed time series 25. The iterated method used here to detect the fixed points of the Poincare’ map is the one proposed in 25. If the Poincare’ map is described by

$$\zeta_{k+1} = f(\zeta_k),$$

the iterative mapping

$$\zeta_{k+1} = f(\zeta_k) + Q(f(\xi_k) - \xi_k) \tag{4}$$

in which

$$Q = (cI - J(\xi_k))(f(\xi_k) - \xi_k)^{-1}$$

converges to the fixed point of the Poincare’ map.

For our model, the fixed point of the map for $\gamma = 0.0185rad$ is found to be

$$\theta_{st}^* = 0.2534$$

$$\dot{\theta}_{st}^* = -0.2453$$

4 Artificial Neural Network

Artificial neural networks are a biologically inspired algorithm with various applications in applied science and engineering problems. The network is basically a simulation of human nervous system composed of some node elements as neurons being related through several interconnection edges as in human nervous system [30]. The composition of several layers with definite number of neurons in each layer and the interconnection between the neurons in layers constructs an artificial neural network [30]. The interconnections have definite weights which collectively determine the output of each neuron and the behavior of the system. First, the ANN should be trained by enough initial information of a set of valid inputs and the associated outputs. Thus the weighted connections associated to the neurons are adjusted so that the ANN learns to perform a definite task for a given set of inputs [30].

In our model ANN is actually trained to model the Poincare' map of the system which cannot be investigated analytically. Once the network learns the dynamics of the system it is prepared to determine the proper control parameter so that the system shows a desirable feature.

In order to control the biped within its chaotic attractor the trajectories should be guided into a fixed point embedded in the region. The fixed point can be of any order of periodicity so that the biped can be transferred from chaotic behavior into a periodic orbit and show flexible behavior which is a major benefit of chaos control. The control parameter which is the toe-off impulse is applied as soon as the states are close to the unstable periodic orbit embedded in the chaotic attractor.

In order to design the controller, an artificial neural network is developed that finds the proper toe impulse to put the trajectory on any desired periodic orbit. The feed-forward network has 23 neurons in the first layer, 4 neurons as the hidden layer and one neuron in the last one. The excitation function of the neurons is sigmoid $f(x) = (1 + e^{-x})^{-1}$ and a back propagation algorithm is used to train the network as used in [24] (Figure 4).

The number of inputs has increased and the controller has been changed compared to [24] to be applicable for the new control parameter and also be able to update the initial conditions as the biped gets new steps. The network is trained to find the proper toe-impulse (P in Eq. 3) for any initial condition and slope angle γ so that the trajectories conform to the unstable periodic orbit.

The inputs of the neural network are the states of the biped $[\theta_{st}, \dot{\theta}_{st}]$ at seven consecutive steps, the desired state of the system at the next step, the six toe-off impulses at every heel strike between the steps and the slope angle of the surface. The states are the input to the first 16 neurons of the first layer, the toe-off impulses are the inputs for the next 6 neurons and the slope angle is the input of the last neuron in the first layer. The desired states are indeed related to the unstable periodic orbit in the chaotic attractor which is detected through the efficient algorithm.

Once the controller is turned on the states of seven consecutive steps at the Poincare' section and the toe-off impulse at jumping moment in each step are stored to be the input for 20 neurons of first layer. The 15th and 16th neurons are the desired fixed point states which is the unstable periodic orbit of the chaotic attractor and the last

one is the value of γ , the slope angle of the surface. Choosing the previous steps as the input is a biologically inspired strategy based on the scientific literature which shows that human locomotion has a neural memory of previous locomotive states 328-[30].

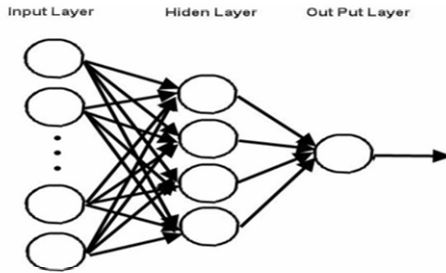


Fig. 4. The schematic of the artificial neural network utilized to control the bipedal robot

5 Simulation Results

The network was trained with a set of 20 data sets and was applied to the model to get a periodic behavior from the biped in the chaotic region.

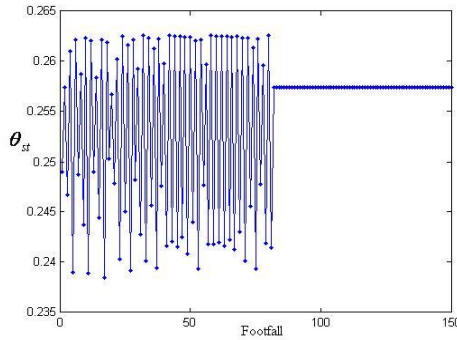


Fig. 5. The unstable periodic orbit of the biped for $\gamma = 0.0185 \text{ rad}$ is stabilized

As shown in Figure 5, once the controller starts to gather the information, the first control signal is generated after 7 steps and immediately the system converges to a periodic orbit.

The network is trained so that can be used to control the biped walking on surfaces with any slope angle. The same neural network controller was applied to the biped walking on surfaces with slope angles of $\gamma = 0.0185 \text{ rad}$, $\gamma = 0.0187 \text{ rad}$ and $\gamma = 0.0189 \text{ rad}$. The designed controller is able to stabilize the bipedal motion on the unstable periodic orbits detected for every slope from the previously mentioned

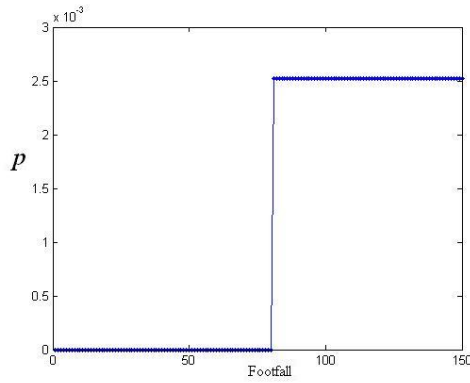


Fig. 6. The variations of control parameter stabilizing the periodic orbit for $\gamma = 0.0185 \text{ rad}$

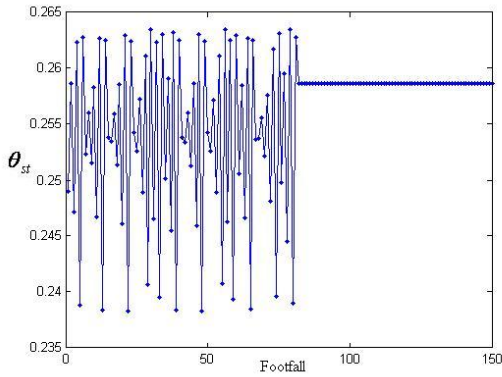


Fig. 7. The unstable periodic orbit of the biped for $\gamma = 0.0187 \text{ rad}$ is stabilized

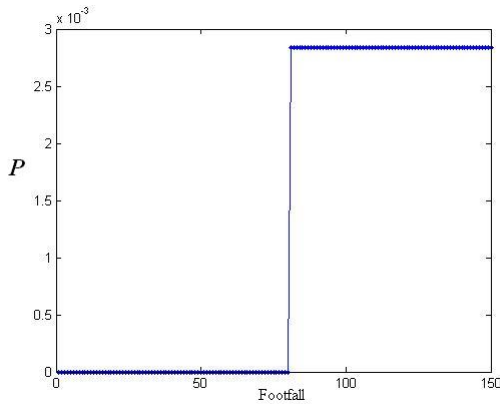


Fig. 8. The variations of control parameter stabilizing the periodic orbit for $\gamma = 0.0187 \text{ rad}$

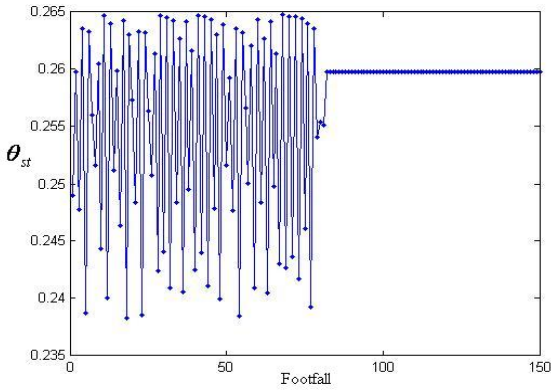


Fig. 9. The unstable periodic orbit of the biped for $\gamma = 0.0189 \text{ rad}$ is stabilized

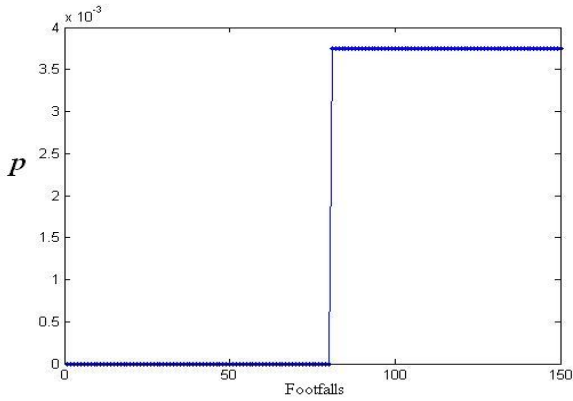


Fig. 10. The variations of control parameter stabilizing the periodic orbit for $\gamma = 0.0189 \text{ rad}$

efficient algorithm. The results are shown in Figures (5)-(10). For each case, the variations of the stance leg angle and the control parameter P at heel strikes are shown.

6 Conclusions

An artificial neural network algorithm was used to control the chaotic behavior in the simplest walking model. The model was proposed as a simple passive walker capable of producing human locomotion. The chaotic behavior was verified and the unstable periodic orbit in the chaotic attractor was detected.

The trained neural network is capable of controlling the biped by toe-impulse signals very rapidly on any arbitrary ramp slope in which chaos occurs.

Since the basin of attraction of chaotic attractor is larger than the one in periodic cycles and the system is more flexible and robust in the chaotic region, a small control effort can produce a flexible robust behavior in the system.

The algorithm is shown to be efficient, robust and flexible and can be applied to more complicated biped robots including the knee masses and an upper body link.

References

1. McGeer, T.: Passive Dynamic Walking. *International Journal of Robotics Research* 9, 62–82 (1990)
2. Clark, J.E., Phillips, S.J.: A Longitudinal Study of Intralimb Coordination in the First Year of Independent Walking: A Dynamical Systems Analysis. *Child Dev.* 64, 1143–1157 (1993)
3. Hausdorff, J.M., Peng, C.K., Ladin, Z., Wei, J.Y., Goldberger, A.L.: Is Walking a Random Walk? Evidence for Long-range Correlations in Stride Interval of Human Gait. *J. Appl. Physiol.* 78, 349–358 (1995)
4. Dingwell, J.B., Cusumano, J.P., Sternad, D., Cavanagh, P.R.: Slower Speeds in Patients with Diabetic Neuropathy Lead to Improved Local Dynamic Stability of Continuous overground Walking. *J. Biomech.* 33, 1269–1277 (2000)
5. Goswami, A., Thuilot, B., Espiau, B.: Compass like Bipedal Robot Part I: Stability and Bifurcation of Passive Gaits, <http://www.inria.fr/RRRT/RR-2996.html>
6. Goswami, A., Thuilot, B., Espiau, B.: A Study of the Passive Gait of a Compass-like Biped Robot: Symmetry and Chaos. *International Journal of Robotic Research* 17, 1282–1301 (1998)
7. Khosravi, B., Yurkovich, S., Hemami, H.: Control of a Four Link Biped in a Back Somersault Maneuver. *IEEE Transactions on Systems, Man, and Cybernetics* 17, 303–325 (1987)
8. Garcia, E., Estremera, J., Gonzales de Santos, P.: A Comparative Study of Stability Margins for Walking Machines. *Robotica* 20, 595–606 (2002)
9. Lum, H.K., Zribi, M., Soh, Y.C.: Planning and Control of a Biped Robot. *International Journal of Engineering Science* 37, 1319–1349 (1999)
10. Saidouni, T., Bessonnet, G.: Generating Globally Optimized Sagittal Gait Cycles of a Biped Robot. *Robotica* 21, 199–210 (2003)
11. Piiroinen, P., Dankowicz, H.: Low-cost Control of Repetitive Gait in Passive Bipedal Walkers. *International Journal of Bifurcation and Chaos* 15, 1959–1973 (2005)
12. Allgood, K.T., Sauer, T.D., Yorke, J.A.: *Chaos: an introduction to dynamical Systems*. Springer, Berlin (1997)
13. Goswami, A., Espiau, B., Keramane, A.: Limit Cycles in a Passive Compass Gait Biped and Passivity-mimicking Control Laws. *Autonomous Robots* 4, 273–286 (1997)
14. Mochon, S., McMahon, T.: Ballistic Walking: An Improved Model. *Mathematical Biosciences* 52, 241–260 (1980)
15. Garcia, M., Chatterjee, A., Ruina, A., Coleman, M.: The Simplest Walking Model: Stability, and Scaling. *ASME Journal of Biomechanical Engineering* 120, 281–288 (1997)
16. Collins, S.H., Ruina, A., Tedrake, R.L., Wisse, M.: Efficient Bipedal Robots Based on Passive-dynamic Walkers. *Science* 307, 1082–1085 (2005)
17. Wisse, M., Schwab, A.L., van der Helm, F.C.T.: Passive Dynamic Walking Model with upper Body. *Robotica* 22, 681–688 (2004)
18. Hurmuzlu, Y.: *Dynamics and Control of Bipedal Robots*. Springer-Verlag Series of Lecture Notes in Control and Information Science, vol. 230, pp. 105–118 (1998)

19. Hurmuzlu, Y., Genot, F., Brogliato, B.: Modeling, Stability and Control of Biped Robots: A General Framework. *Automatica* 40, 1647–1664 (2004)
20. Wisse, M.: Essentials of Dynamic Walking; Analysis and Design of Two-legged Robots. Ph.D. thesis, T.U. Delft (2004)
21. Starrett, J., Tagg, R.: Control of a Chaotic Parametrically Driven Pendulum. *Phys. Rev. Lett.* 74, 1974–1977 (1995)
22. Ott, E., Grebogi, C., Yorke, J.A.: Controlling chaos. *Phys. Rev. Lett.* 64, 1196–1199 (1990)
23. Suzuki, S.: Passive Walking Towards Running. *Mathematical and Computer Modelling of Dynamical Systems* 11, 371–395 (2005)
24. Kurz, M.J., Stergiou, N.: An Artificial Neural Network that Utilizes Hip Joint Actuators to Control Bifurcations and Chaos in a Passive Dynamic Hipedal Walking Model. *Biol. Cybern.* 93, 213–221 (2005)
25. Bu, S., Wang, B.-H., Jiang, P.-Q.: Detecting Unstable Periodic Orbits in Chaotic Systems by Using an Efficient Algorithm. *Chaos, Solitons and Fractals* 22, 237–241 (2004)
26. Buhl, M., Kennel, M.B.: Globally Enumerating Unstable Periodic Orbit Theory for Observed Data Using Symbolic Dynamics. *Chaos* 17, 033102 (2007)
27. Mark, D.A.: *Analytical Dynamics: Theory and Applications*. Kluwer Academic/Plenum Publishers, New York (2005)
28. Hausdorff, J.M., Mitchell, S.L., Firtion, R., Peng, C.K., Cudkowicz, M.E., Wei, J.Y., Goldberger, A.L.: Altered Fractal Dynamics of Gait: Reduced Stride-interval Correlations with Aging and Huntington’s disease. *J. Appl. Physiol.* 82, 262–269 (1997)
29. Hausdorff, J.M., Zemani, L., Peng, C.K., Goldberger, A.L.: Maturation of Gait Dynamics: Stride-to-stride Variability and Its Temporal Organization in Children. *J. Appl. Physiol.* 86, 1040–1047 (1999)
30. Martin, T.H., Howard, B.D., Beale, M.: *Neural Network Design*. PWS Publishing Company, Boston (2002)

Reinforcement Learning Control of a Real Mobile Robot Using Approximate Policy Iteration^{*}

Pengcheng Zhang, Xin Xu, Chunming Liu, and Qiping Yuan

Institute of Automation, National University of Defense Technology,
410073 Changsha, China
xuxin_mail@263.net

Abstract. Machine learning for mobile robots has attracted lots of research interests in recent years. However, there are still many challenges to apply learning techniques in real mobile robots, e.g., generalization in continuous spaces, learning efficiency and convergence, etc. In this paper, a reinforcement learning path-following control strategy based on approximate policy iteration (API) is developed for a real mobile robot. It has some advantages such as optimized control policies can be obtained without much *a priori* knowledge on dynamic models of mobile robot, etc. Two kinds of API-based control method, i.e., API with linear approximation and API with kernel machines, are implemented in the path following control task and the efficiency of the proposed control strategy is illustrated in the experimental studies on the real mobile robot based on the Pioneer3-AT platform. Experimental results verify that the API-based learning controller has better convergence and path following accuracy compared to conventional PD control methods. Finally, the learning control performance of the two API methods is also evaluated and compared.

Keywords: Mobile robots, Approximate policy iteration, Reinforcement learning, Path following, Approximate dynamic programming.

1 Introduction

In the last decade, the control method of wheeled mobile robots (WMR) has become a more and more important research area in artificial intelligence and control theory. According to the description in [1], the wheeled mobile robots are categorized into five large groups which include unicycle, two-wheeled robot, three-wheeled robot, four-wheeled (car-like) robot and tractor-trailer system. Here, the research work will be focused on a two-wheeled mobile robot.

Path following control is a typical example of wheeled mobile robots which has been studied for decades since 1980s [2] [3] and lots of results have been achieved. Until now, much research work has been devoted to the construction of lateral controllers for mobile robots [5] [6], which include PID methods, fuzzy controllers, slide model

^{*} Supported by the National Natural Science Foundation of China (NSFC) under Grants 60774076, 90820302, the Fok Ying Tung Education Foundation under Grant No.114005, and the Natural Science Foundation of Hunan Province under Grant 07JJ3122.

controllers, etc. However, these methods can not realize performance optimization of mobile robots with model uncertainties, especially in some unknown environments.

Reinforcement learning (RL) assumes that the robot can be modeled as a Markov decision process (MDP) and the learning agent interacts with an initially unknown environment and modifies its action policies to maximize its cumulative payoff [8], [9]. Consequently, there is no need to program the desired knowledge to robot by human in a RL-based control strategy [7], [10]. Thus, combining RL-based control strategy with path following control has attracted more and more interests in recent years. However, two main obstacles still exist for the wider application of RL in real mobile robot control. One problem is that the local convergence of existing gradient-based RL algorithms always hinders the achievement of optimal or near-optimal policies. The other is that many RL algorithms rely heavily on manually selected approximation structures. In order to solve these problems, a path following control strategy using the recently developed approximate policy iteration (API) method, which includes least-squares policy iteration (LSPI) [14] and kernel LSPI (KSLPI) [11], is studied and evaluated in a real mobile robot in this paper. The API method can discover an optimal or near-optimal policy for an MDP by generating a sequence of improved policies and the approximations are realized in two aspects, one is the representation of the value function, the other is the representation of the policy. Two kinds of API-based control methods, i.e., LSPI and KLSPI, are implemented in the path following control task and the efficiency of the proposed control strategy is illustrated in the experimental studies on the real mobile robot based on the Pioneer3-AT platform. Experimental results verify that the API-based learning controller has better convergence and path following accuracy compared to conventional PD control methods. Furthermore, the performance of KLSPI-based learning control and LSPI-based learning control is also evaluated and compared.

This paper is organized as follows. In section 2, the control task and the mobile robot will be introduced. In section 3, the API-based learning control strategy is presented. The experimental results will be performed in section 4. Section 5 concludes the paper.

2 Problem Description

In this paper, the task for the mobile robot is a geometric path following. The robot must sense its position with respect to the desired path and return to the path if it is off course. By using the sensors on the wheeled mobile robot, the lateral controller calculates the error and determines how to turn the steering wheels to follow the desired path. The path in the lab contains a white line on a black surface which the robot is to follow.

A real mobile robot is used in this research and it is P3-AT wheeled mobile robot. As a differentially-driven platform, P3-AT mobile robot supports a combined control of linear velocity and angular velocity or a differential control of the wheels. The position and velocity can be measured by the direct drive motor in which a high-resolution optical encoder installed.

The simplified geometry model of P3-AT mobile robot can be described by Fig.1.

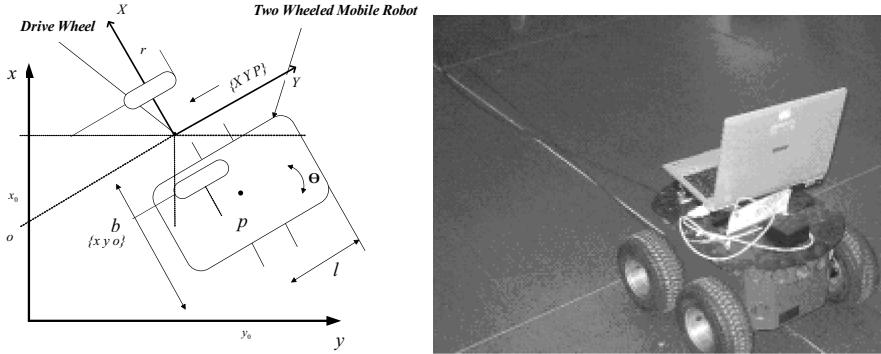


Fig. 1. The geometry model and image of P3-AT mobile robot

Here, vector $(x \ y \ \theta)^T$ is used to describe the pose of point P which represents the center of the robot in Cartesian coordinate $\{x \ y \ o\}$. Where $(x_0, \ y_0)$ defines the x -coordinate value and y -coordinate value of point P in Cartesian coordinate, θ is the angle between the coordinate of WMR $\{X \ Y \ P\}$ and the Cartesian coordinate $\{x \ y \ o\}$.

The desired path is set to be a straightaway in the x -coordinate direction and the desired position of WMR is set as $(x_d, \ y_d)$. The learning control task is to minimize the accumulative error between the actual path and the desired path when there is a large initial position error and the dynamic model of WMR is assumed to be unknown. The objective function can be described by following equation

$$J = \sum_{t=0}^T e(t) \tag{1}$$

where $e(t)$ stands for the path following error and J is the accumulative error calculated from the start point to the end point.

3 An API-Based Learning Control Framework

According to the control task of the mobile robot discussed in section 2, a framework of the learning control method based on API is proposed in this section. The API method is embedded into the design of the learning control system and the learning optimizing problem is formulated as a Markov decision process (MDP).

3.1 API-Based Optimization for Path-following Control System

As a class of actor-critic learning algorithms, API method can be viewed as a hybrid of value function approximate and policy learning. It consists of a policy evaluation module and a policy improvement module. Policy evaluation is to evaluate the state-action value function in fixed policies, and policy improvement is to gain the greedy policies based on the state-action value function evaluation. In Fig. 2, the API-based learning controller is integrated with a conventional proportional-derivative (PD) path following controller.

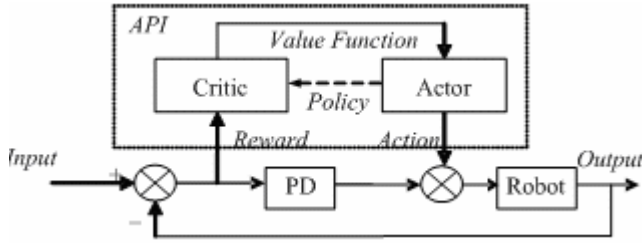


Fig. 2. The API-based learning control system

In API, policy evaluation usually estimate the value function by using TD learning algorithms without any model information, the TD error is calculated by comparing an expected value from an actually obtained reward which is determined by the following error between desired and actual path in this paper. The learning process of API-based control policy can be described by the following flow chart:

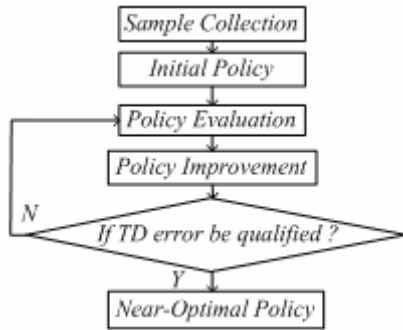


Fig. 3. The learning process of API-based control policy

Fig. 3 shows that based on the evaluation of policies at last iteration, the actor improves the policy and the iteration process is repeated until there is no change between the newly produced policies. Consequently, once the TD learning algorithms are well performed, the optimal or near-optimal control policy can be obtained, which will be discussed in subsection 3.4.

By approximating the state value function $Q^\pi(s, a)$ of policy π , the corresponding policy is obtained as:

$$\pi_{t+1}(s) = \arg \max_a \tilde{Q}^{\pi_t}(s, a) \tag{2}$$

where π_t is the policy at t th iteration, and π_{t+1} is the improved policy based on the estimated action value function.

3.2 The MDP Formulation of Controller Optimization

In this section, the controller optimization of the mobile robot is formulated as a Markov decision process (MDP) and the expected result is realized by minimizing the following error between actual and desired path. The MDP is denoted as a tuple $\{S, A, P, R\}$, where S is the state space, A is the action space, P is the state transition probability, and R is the reward function.

According to the learning control problem of mobile robot studied in this paper, the state space in MDP can be defined as $s(t) = [e_x, e_y, e_\theta]$ in which the three-dimensional vector represents the error between actual and desired position of robot. $A = \{a_1, a_2, \dots, a_n\}$ is used to represent the infinite action sequence and the PD coefficients are optimized by API-based learning strategy, then the action space of MDP can be defined as a series of PD coefficients which is expressed by $a(t) \in [(k_{p1}, k_{d1}), (k_{p2}, k_{d2}), \dots, (k_{pn}, k_{dn})]$. Based on the defined reward function, the lateral control performances of the mobile robot are optimized by choosing the appropriate actions. To optimize the path following performance of wheeled mobile robot, the reward function and the target function can be described as follows:

$$r_t = c|e_t| \tag{3}$$

$$J = \sum_{t=0}^T \gamma^t r_t \tag{4}$$

where the e_t is the error between desired and actual path of mobile robot, c is a negative constant, γ is the discount factor. The temporal differences are defined as the differences between two successive estimations and it can be obtained as the following form

$$\delta_t = r_t + \gamma Q_{t+1}(s, a) - Q_t(s, a) \tag{5}$$

3.3 The API-based Learning Control Methods

Based on the work of least squares temporal difference (LSTD) learning algorithm in [13], least squares policy iteration (LSPI) was proposed in [14], it performs better properties in convergence, stability, and sample complexity than previous RL algorithms. In order to solve the automation feature selection and improve the convergence performance of LSPI, Xu etc., applied kernel methods to LSTD algorithm and proposed KLSPI algorithm in [11]. The value functions can be described as following equations

$$Q(x) = \phi^T(x)W = \sum_{j=1}^n \phi_j(x)w_j \tag{6}$$

$$Q(x, a) = \sum_{i=1}^l \alpha_i k(s, s_i) \tag{7}$$

where equation (6) and equation (7) are the value function of LSPI and KLSPI algorithm respectively, $\phi(x) = [\phi_1(x), \phi_2(x), \dots, \phi_n(x)]^T$ is a vector of basis functions, x is an observation state in the trajectory $\{x_t | t = 0, 1, 2, \dots; x_t \in X\}$ generated by a Markov chain, $k(s, s_i)$ is the kernel function, s and s_i are the combined features of state-action

pairs (x, a) and (x_t, a_t) , respectively, $\alpha_i (i = 1, 2, \dots, t)$ are the coefficients, W is the weight vector which is used to calculate the value function and can be obtained by following equation

$$E_0[A(X_t)]W^* - E_0[b(X_t)] = 0 \tag{8}$$

where $X_t = (x_t, x_{t+1}, z_t)$ ($t = 1, 2, \dots$) from a Markov process. $E_0[\cdot]$ stands for the expectation with respect to the unique invariant distribution of $\{X_t\}$. $A(X_t)$ and $b(X_t)$ are defined as follows

$$A(X_t) = \bar{z}_t(\phi^T(x_t) - \gamma\phi^T(x_{t+1})) \tag{9}$$

$$b(X_t) = \bar{z}_t r_t \tag{10}$$

$$A_t = A_{t-1} + \bar{k}(s_t)[\bar{k}^T(s_t) - \gamma\bar{k}^T(s_{t+1})] \tag{11}$$

$$b_t = b_{t-1} + \bar{k}(s_t)r_t \tag{12}$$

where equation (9) and equation (10) are the definitions in LSPI algorithm, z_t is the eligibility trace vector, equation (11) and equation (12) are the definitions in KLSPI algorithm.

According to the above discussions, the framework of LSPI and KLSPI based learning controller can be described as follows

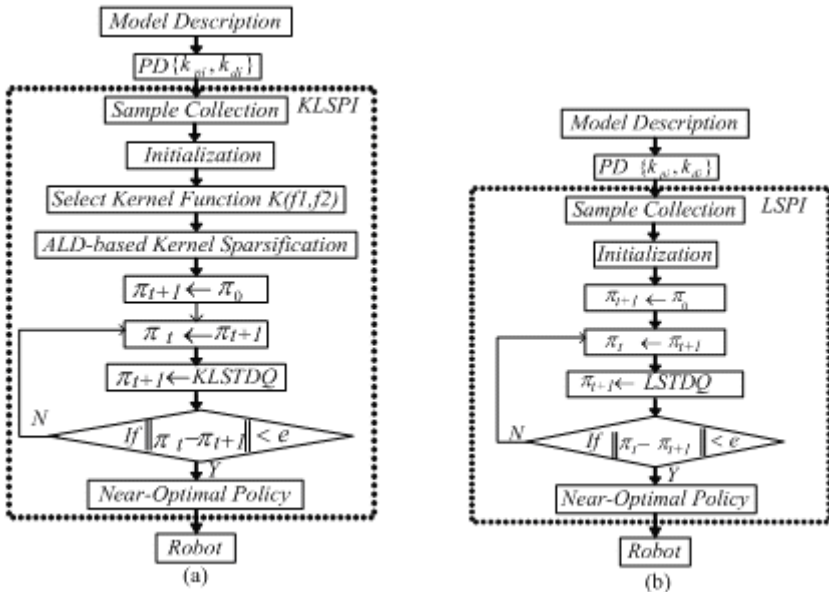


Fig. 4. The KLSPI-based and LSPI-based learning control method are shown in (a) and (b)

π_0 and π_t are the initial policy and the current policy, respectively, π_{t+1} is the updated policy based on the API method, e is the positive scalar between current control policy and optimized policy based on LSTDQ or KLSTDQ algorithm. Aiming to keep

the sparsity and improve the generalization ability of KLSTDQ algorithm, a kernel sparsification procedure using approximate linear dependency (ALD) is performed. Some related work is discussed in [11].

3.4 Performance Analysis

In order to analyze the convergence of the API-based control methods, a genetic theorem which was proposed by Bertsekas and Tsitsiklis (1996) [16] is introduced. It proves that the API methods can converge to the optimal or near-optimal policy when the error in policy evaluation and projection and the error in policy improvement and projection are bounded [17]. Thus, the performance of the optimized policies can be described by the following equation

$$\limsup_{m \rightarrow \infty} \|Q^{\pi_m}(s, a) - Q^*(s, a)\|_{\infty} \leq \frac{\delta + 2\gamma\varepsilon}{(1 - \gamma)^2} \quad (13)$$

Where ε and δ are positive scalars that bound the error in all approximations to value functions and policies respectively, γ is the discount factor, Q^{π_m} and Q^* are corresponding sequence approximate value functions and optimal value function of policies π_m ($m = 1, 2, 3, \dots$) and optimal policy π^* respectively.

4 Controller Implementation and Evaluation

In the following, the performances of API-based learning control methods are tested on the real mobile robot platform Pioneer 3-AT (P3-AT).

4.1 Controller Design for Path Following of WMR

According to the discussions in section 2, the linear velocity is set as $v_p = 0.3m/s$ at the beginning and the mobile robot is set to run towards the positive direction of x -coordinate. The desired path is set as skip from straight line $y = 0$ to straight line $y = 1$ and the mobile robot is expected to following the desired path with a stationary linear velocity.

The PD controller is selected to the lateral control and the following equations can be obtained

$$\omega_p(t) = k_p e(t) + k_d \dot{e}(t) \quad (14)$$

where the error and the derivative of error between desired and actual path are given by

$$\begin{cases} e(t) = 1.0 - y \\ \dot{e}(t) = -v_p \sin \theta \end{cases} \quad (15)$$

4.2 Sample Collection

During the experiment, the initial position of mobile robot is set as $x = -2m$ and run towards the positive direction of x -coordinate. Thus, the linear velocity v_p will be $0.3m/s$ when the mobile robot arrives at $x = 0m$ and then the lateral control begin.

The coefficients of the controller are chosen from the PD parameters by a random strategy. Every selected set of parameters last $0.5s$ in the lateral control process and the position information will be recorded. The mobile robot keeps running for $4m$ in the x -coordinate and a new set of parameters is selected every $0.5s$. When the robot arrives at $x = 4m$, the first period of sample collection is completed and $y+1.0$ is set to be the new value in y -coordinate direction. Then, the second period of sample collection will begin and the data sampling of the mobile robot will be completed by repeating this control process.

There are totally 800 data sampling periods in our experiment, 21539 sets of position data serial $\{(x_i, y_i, \theta_i)\}$ and motion serial $\{a_i\}$ were collected, where $i = 1, 2, \dots, 21539$, $a_i \in \{(k_{i1}, k_{d1}), (k_{i2}, k_{d2}), (k_{i3}, k_{d3})\}$. By filtering the original data, 11244 samples with MDP characteristics will be obtained.

4.3 Comparisons of Conventional Control Methods and API-Based Methods

The experiment on real mobile robot is performed by the following steps:

Firstly, by applying the method aforementioned, 11244 samples are collected in the experiment and the PD parameters can be chosen as $\{k_{pi}, k_{di}\} = \{(0.6, 0.9), (0.4, 1.2), (0.2, 1.5)\}$.

Secondly, the LSPI-based learning control method is combined with PD controllers, the initial training samples are generated by a random control policy in which the discount factor is chosen as $\gamma = 0.9$ and the approximate error is set as $e = 10^{-5}$.

Thirdly, the KLSPI-based control method is implemented to the path following control task. Based on the construction of KLSPI method, the state-action value function and kernel function are chosen as follows.

$$feature(state, action) = \begin{cases} (state(1) \quad state(2) \quad 0 \quad 0 \quad 0 \quad 0) \text{ if } action = 1 \\ (0 \quad 0 \quad state(1) \quad state(2) \quad 0 \quad 0) \text{ if } action = 2 \\ (0 \quad 0 \quad 0 \quad 0 \quad state(1) \quad state(2)) \text{ if } action = 3 \end{cases} \quad (16)$$

$$Kernel(f1, f2) = e^{-\left(\sum_{i=1}^6 |f1(i) - f2(i)|\right) / 0.5} \quad (17)$$

According to equation (17), the radius basis function (RBF) kernel function is used and the width is set as $\sigma = 0.5$, where $f1 = feature(state1, action1)$ and $f2 = feature(state2, action2)$. As the threshold parameter for kernel sparsification is set as $\mu = 0.2$, a feature vector with 257 dimensions will be obtained. The experimental results obtained by different controllers are shown as follows.

Table 1. Path following performances under API-based controllers versus PD controllers

Lateral Controllers		Performance		
		Convergence Time	Over Shoot	Ascend Time
PD (k_{pi}, k_{di})	(0.6, 0.9)	25s	0.35	4.3s
	(0.4, 1.2)	18.6s	0.08	7s
	(0.2, 1.5)	24s	0	24s
LSPI-based		17.8s	0.17	4.8s
KLSPI-based		13s	0.02	6.5s

From Table 1, it is illustrated that API-based path following control methods can converge to the near-optimal policy with a much better performance than the conventional PD controllers. In addition, the feature selection is automatically by using the ALD kernel sparsification approach in KLSPI-based control method. The path following images of real mobile robot are shown as follows.

4.4 Comparisons of LSPI-Based and KLSPI-Based Learning Control Method

Fig. 5 shows the value functions of LSPI-based and KLSPI-based learning control method after convergence.

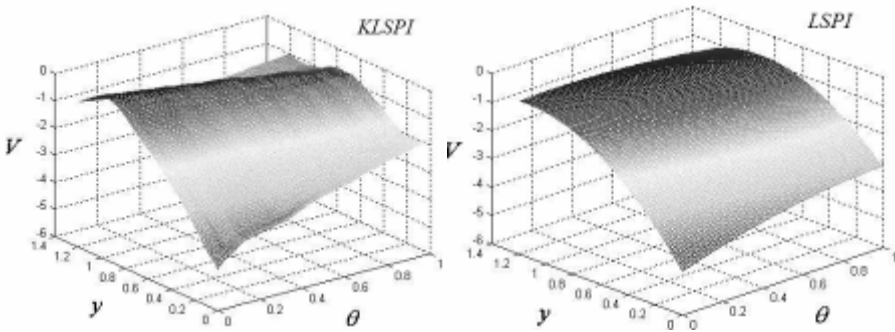


Fig. 5. Value functions of API-based learning control method

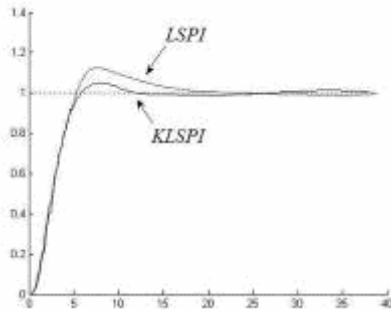


Fig. 6. Path following curves of API-based learning control methods

From Fig. 6, it is illustrated that the convergence and optimality can be guaranteed based on the KLSPI method and the performance of KLSPI can be better than LSPI in terms of higher precision and less iterations in the path following control task.

5 Conclusions

In this paper, an API-based control strategy is implemented in the path following control task of a P3-AT mobile robot. We have described the nonlinear system of a real differential-drive WMR, focusing on the optimization problem of its learning control architecture. The experimental results verify that much better path-following performance than conventional controllers can be obtained by applying the proposed control strategy.

The API-based control strategy makes mainly three improvements to the conventional path following controllers. Firstly, the near-optimal policy can be obtained without much a priori knowledge on dynamic models of mobile robot. Secondly, the performance of the learning control system with better convergence rate and generalization ability can be well guaranteed adaptively by incorporating conventional controller as initial policies. Thirdly, the learning control efficiency has been greatly improved due to the automatic feature selection in which the ALD-based kernel sparsification method is implemented. Although the results are very encouraging, the applications of API method in more complicate control problems such as continuous action spaces are to be studied in future.

References

1. Campion, G.: Structural Properties and Classification of Dynamic Models of Wheeled Mobile Robots. *IEEE Trans. on Robotics and Automation* 12, 47–62 (1996)
2. Alexander, J.C., Brooks, J.H.: On the Kinematics of Wheeled Mobile Robots. *Int. J. of Robotics Research* 8, 15–27 (1989)
3. Chiacchio, P.: Exploiting Redundancy in Minimum-time Path Following Robot Control. In: *American Control Conference* (1982)
4. Sarkar, N., Gen, V.: Dynamic Path Following: A New Control Algorithm for Mobile Robots. In: *32nd Conference on Decision and Control*, pp. 2670–2675. IEEE Press, New York (1993)
5. Coelho, P., Nunes, U.: Path Following Control of Mobile Robots in Presence of Uncertainties. *IEEE Transaction on Robotics* 21, 252–261 (2005)
6. Brooks, R.: A Hardware Retargetable Distributed Layered Architecture for Mobile Robot Control. In: *IEEE International Conference on Robotics and Automation*, pp. 106–110. IEEE Press, New York (1987)
7. Chen, C.L., Chen, C.H.: Reinforcement Learning for Mobile Robot from Reaction to Deliberation. *Journal of Systems Engineering and Electronic* 16, 611–617 (2005)
8. Sutton, R., Barto, A.: *Reinforcement Learning, an Introduction*. MIT Press, Cambridge (1998)
9. Kaelbling, L.P., Littman, M.L., Moore, A.W.: Reinforcement Learning: A Survey. *J. Artif. Intell. Res.* 4, 237–285 (1996)

10. Smart, W.D., Kaelbling, L.P.: Effective Reinforcement Learning for Mobile Robots. In: IEEE International Conference on Robotics and Automation, pp. 3404–3410. IEEE Press, New York (2002)
11. Xu, X., Hu, D.W., Lu, X.C.: Kernel-Based Least Squares Policy Iteration for Reinforcement Learning. IEEE Transaction on Neural Networks 18, 973–992 (2007)
12. Canudas, C., Sordalen, O.J.: Exponential Stabilization of Mobile Robots with Non-holonomic Constraints. IEEE Transactions on Automatic Control 33, 672–677 (1992)
13. Boyan, J.: Technical Update: Least-squares Temporal Difference Learning. Mach. Learn. 49, 233–246 (2002)
14. Lagoudakis, M.G., Parr, R.: Least-squares Policy Iteration. J. Mach. Learn. Res. 4, 1107–1149 (2003)
15. Engel, Y., Mannor, S., Meir, R.: The Kernel Recursive Least-squares Algorithm. IEEE Trans. Signal Process. 52, 2275–2285 (2004)
16. Bertsekas, D.P., Tsitsiklis, J.N.: Neuro-Dynamic Programming. Athena Scientific, Belmont (1996)
17. Lagoudakis, M.G., Parr, R.: Least-squares Policy Iteration. J. Mach. Learn. Res. 4, 1107–1149 (2003)

A Simple Neural Network for Enhancement of Image Acuity by Fixational Instability

Daqing Yi¹, Ping Jiang^{1,2}, and Jin Zhu¹

¹ Department of Information and Control Engineering,
Tongji University, Shanghai 200092, China

² Department of Computing, The University of Bradford,
Bradford BD7 1DP, UK
Dqyi11@gmail.com

Abstract. Inspired by biological findings, this paper proposes a neural network model for achieving higher image acuity by introducing random eye movement. Statistical analysis and comparison study of the image quality in the presence and absence of random eye movement are carried out using the model. It is revealed that, as a noise source to a stationary image, the random eye movement can contribute to overcome the inherent resolution limits of photoreceptors and enhance sharpness of images by temporal statistics of firing neurons. Super-resolution and prominent edges can thus be achieved, with superior visual acuity to the absence of eye-movement. The acuity enhancement is in fact a trade-off between bias and variance and is related to the distribution of visual stimuli and eye-movement patterns. The simulations illustrate its effect on enhancement of image acuity.

Keywords: Super-resolution; Image acuity; Eye-movement; Statistical neural networks.

1 Introduction

In 1738, Jurin found that our eyes are never still, trembling exists even when fixation, which means this kind of eye-movement occurs involuntarily when people gazes at a particular object. He named this phenomenon as the “trembling of the eye”. Since then, a variety of techniques for recording eye-movement have been developed. But till 1952, R. Ditchburn and B. Ginsborg analyzed the existence of involuntary eye-movements during fixation in a scientific way [1]. At present, scientists agree on the occurrence of three main types of eye-movements during visual fixation in humans, which are tremor, drifts and microsaccades respectively [12].

Since being discovered [2], the functional role of the fixational eye-movement has been debated for years. Techniques have been improved in exact observing on the pattern of trembling [3][5]. Parametric analysis of eye tremor has been reported in [11]. In a laboratory condition called “retinal stabilization”[12], the visual percept rapidly fades out due to the absence of eye ball jitter and may even completely disappear under certain conditions, which is known as “Troxtler’s effect”. Initially, as a visual system is

sensitive to moving objects, the reason for fixational eye-movement is explained as preventing neural adaption, which fits the “Troxler’s effect”. When gazed at an unchanged environment, the neural adaption will occur and will drive the vision fading out [13]. However, more and more observations show that it cannot explain all phenomena caused by fixational eye-movement. In the current consensus, fixational eye-movement occurring involuntarily when people gaze at a particular object contributes to maintain visibility. Paper [4] gave a comparison of motion perception between existence and non-existence of involuntary visual vibration, and came to a conclusion that undetected trembling in static eyes contributed to the rise of the perception of illusory motion. Paper [6] discussed the relationship between the “fixation eye movement” and the retinal information processing mechanism, which gave an explanation on the adaptation of vision and the disappearance of image in a static retina. From an engineering perspective, paper [7] analyzed how periodic visual vibrations contributed to resolution enhancement and implemented a visual sensing microsystem taking advantage of this principle. Paper [8] considered the eye micro-movement contributed to stimulus detection, which was beyond the Nyquist limit in the peripheral retina. Paper [9] also held an idea that eye-tremor actually improved visual acuity under the stochastic resonance in visual cortical neurons. Paper [13] pointed out that fixational eye-movement may help to disambiguate latency and brightness during visual perception, which allows latency to be used for visual discriminations. From experiments, the fixational eye-movement was found to increase the retinal activity [12] by moving their receptive fields over stationary stimulus to form long and tight spike bursts.

Inspired by the biological findings, this paper proposes a simple statistical neural network for enhancement of image acuity by introducing eye movement. Statistical analysis reveals that the light stimulus to a photoreceptor follows a trapezoidal distribution with zero mean as the result of uniform random eye movement. The statistical characteristics of neuron codes fired by the stimulus are then derived. It shows that the eye movement can contribute to enhancement of image acuity subject to a proper eye movement pattern, which is a tradeoff between bias and statistical period, i.e. perception accuracy and response time. It provides a design reference for development of engineering visual sensors to break their perception limit by introducing active vibration. Simulations are carried out to demonstrate the capability of random eye movement for achieving the effect of super-resolution and prominent edges.

2 A Statistical Visual Neural Model

The front end of a visual system consists of an ocular system and a retina. The ocular system includes cornea, lens, iris and so on. The retina is a thin layer of neural cells that lines the back of an eyeball, which responds light rays for building an image of the visual world. The retina contains photoreceptor cells, which are rods and cones that respond to the light. As light with different intensity reaches a photoreceptor, the photoreceptor cell produces different membrane potential as a response. Then this form of neural impulse is further converted into neural codes in retina ganglion cell for imaging and interpretation in the brain [12].

This light encoding mechanism of a photoreceptor in response to light stimulus I can be mimicked by an inter-inhibited neural network consisting of McCulloch-Pitts

Neurons. A neuron in the first layer can be activated only if the light stimulus I passes its threshold, θ_i . Then the activated neurons attempt to fire the counterpart neurons in the second layer but inhibit firing of those neurons with lower thresholds. The proposed visual neural model can be expressed as:

$$y_i = \begin{cases} 1 & (I > \theta_i) \text{ and } (I < \theta_j \text{ for all } j > i) \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where $i = -M \dots M$ and $\theta_{-M} < \theta_{-M+1} < \dots < \theta_M$.

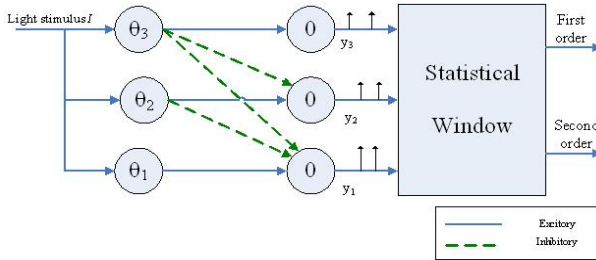


Fig. 1. The structure of the neural network ($\theta_3 > \theta_2 > \theta_1$)

This neural network encodes a light intensity to an activated y_i . A fired neuron i , represented by a neural code of $[I]$, indicates light stimulus I perceived by this photoreceptor satisfies $\theta_i < I < \theta_{i+1}$. Thus different stimulus provokes firing of different neuron to reflect its intensity range.

For simplifying further analysis, we assume evenly separated thresholds around 0 with an increment of R as shown in Fig.2, i.e.

$$\{\theta_{-M}, \dots, \theta_{-1}, \theta_0, \theta_1, \theta_2, \dots, \theta_M\} = \left\{ \frac{-2M-1}{2}R, \dots, -\frac{3R}{2}, \frac{R}{2}, \frac{R}{2}, \frac{3R}{2}, \dots, \frac{2M-1}{2}R \right\} \tag{2}$$

In the case of still fixation without any eye-movement, the output of the neural code is a deterministic value in response to stimulus with a possible quantization error, which corresponds to a neural code having lower resolution.

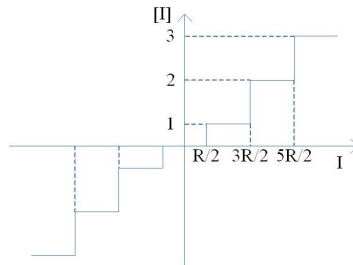


Fig. 2. The deterministic neural code in response to I

By introducing random eye movement, the adjacent neurons can be fired and temporal spike sequences are generated at those y_i 's. An example of the generated temporal spike sequences due to eye movement is shown in Fig.3. It is expected that the first order temporal statistics of spikes is able to recover the stimulus with higher accuracy and the second order statistics is able to sharpen the edges in an image.

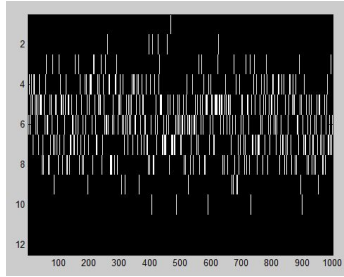


Fig. 3. The neural spike trains under the activation of random jittering

3 Eye-Jittering and Neuron Firing

The neural responses due to eye-movement are similar to the responses if a visual stimulus moves over the stationary receptive field of a visual neuron [13]. Assume the light intensity of an image is $I(x, y)$. The changes on the intensity due to eye-movement can be approximated by the optical flow constraint equation:

$$\Delta I(x, y) = -\Delta x \cdot \partial I / \partial x - \Delta y \cdot \partial I / \partial y = a(x, y) \cdot \Delta x + b(x, y) \cdot \Delta y \tag{3}$$

where Δx and Δy are the eye-movement in the x and y directions, and $a(x, y) = -\partial I / \partial x$ and $b(x, y) = -\partial I / \partial y$ are spatial intensity gradients at (x, y) ; $\Delta I(x, y)$ is the intensity changes as the consequence of the eye-movement. The movements in the x and y directions are assumed to be independent with zero means, i.e. $Cov(\Delta x, \Delta y) = 0$, $E(v_x) = E(\Delta x) = 0$ and $E(v_y) = E(\Delta y) = 0$. Hence $E(\Delta I) = 0$ due to (3). Further assume Δx and Δy are random movements with a uniform density function in a range of $[-K, K]$, i.e.

$$P(\Delta x) = P(\Delta y) = \begin{cases} \frac{1}{2K} & x, y \in [-K, K] \\ 0 & \text{others} \end{cases} \tag{4}$$

From (3), the probability density function of the consequent intensity change can be obtained by convolution between $P(\Delta x)$ and $P(\Delta y)$.

$$f(\Delta I) = \frac{1}{b} \int_{-\infty}^{+\infty} P(\Delta x) P\left(\frac{a}{b}(\Delta x - \frac{\Delta I}{a})\right) d\Delta x \tag{5}$$

Therefore, the possibility density function of intensity variation caused by the uniformly distributed eye-jittering is in a form of trapezoid. An observed area with higher intensity gradient stimulates a wider range of intensity changes to a photoreceptor, which is equal to $2(a+b)K$. When $a = b$, the trapezoid degenerates to a triangle. Taking into account this movement caused intensity variation, the actual light intensity perceived by the photoreceptor at (x,y) becomes $X(x,y) = I(x,y) + \Delta I$.

Thus when this vibrating stimulus reaches a photoreceptor cell, several neurons around the neuron [I] in the neural network in Fig.1 would be fired, denoted as $[I + \Delta I(t)]$. It is expected that statistics of the firing neurons can reveal light stimulus with higher fidelity than the original resolution.

4 Statistical Analysis of Neural Codes

As stated in [13], difference in contrast and salience is represented as the difference in visual responses in the brain, and fixational eye-movement seems to represent a mechanism for enforcing and refreshing information coming from stationary visual stimuli by spatial summation and temporal summation. Motivated by this, the statistics of the neural codes, in the form of interspike interval, was introduced into the neural model in Fig.1.

Assume that samples of the neural codes for T periods (statistical window) are denoted as $[X_t], t=1..T$. The temporal average of T periods can be used as an estimation of the original stimulus value:

$$(\bar{X}) = \sum_{t=1}^T [X_t]R / T \tag{6}$$

Because the eye-movement is uncorrelated, the mean and the variance of (\bar{X}) can be known straightforwardly as

$$E(\bar{X}) = E([X_t]R) \text{ and } Var(\bar{X}) = Var([X_t]R)/T \tag{7}$$

From (1) and (5), we know that $Var([X_t]R) < [(a+b)K/R + 1]^2 R^2$, which is bounded. The variance of (\bar{X}) can be decreased to any value lower than the photoreceptor resolution R by selecting a proper T . Therefore, incorporating eye movement and firing codes statistics provides capability to reduce perception fluctuation. Higher variance of (\bar{X}) , which may be caused by higher intensity gradients or larger eye movement, requires a longer period of statistics, i.e. a slower response, to have a lower variance. However, less fluctuation of the neural estimation does not mean a high visual fidelity. Bias of the statistic estimation, $\Delta = E(\bar{X}) - I$, is another key factor. From (7), $E([X_t]R)$ needs to be derived in order to examine the bias. Because $I = [I]R + (I - [I]R)$ with $(I - [I]R) \in (-R/2, R/2)$, we can suppose $I \in (-R/2, R/2)$ in the following analysis but without losing generality, only subject to a constant shift of $[I]R$.

Assume that, for the thresholds defined in (2) and the intensity distribution in (5), $(a+b)K = (mR + \Delta_2)$ and $|a-b|K = (nR + \Delta_1)$, where mR is the central value of the m th neuron which is fired by intensity $(a+b)K$, i.e. $\theta_m < (a+b)K < \theta_{m+1}$, and the deviation from the center is denoted by Δ_2 ; nR and Δ_1 have a similar meaning for $|a-b|K$. Therefore, $\Delta_1, \Delta_2 \in (-\frac{1}{2}R, \frac{1}{2}R)$.

Then probability density function (5) of the random stimulus can be rewritten as:

$$f(x) = \begin{cases} 0 & x \in (-\infty, -(mR + \Delta_2)] \\ \frac{x + (mR + \Delta_2)}{((m+n)R + \Delta_1 + \Delta_2)((m-n)R + \Delta_2 - \Delta_1)} & x \in (-(mR + \Delta_2), -(nR + \Delta_1)] \\ \frac{1}{((m+n)R + \Delta_1 + \Delta_2)} & x \in (-(nR + \Delta_1), (nR + \Delta_1)] \\ \frac{-x + (mR + \Delta_2)}{((m+n)R + \Delta_1 + \Delta_2)((m-n)R + \Delta_2 - \Delta_1)} & x \in ((nR + \Delta_1), (mR + \Delta_2)] \\ 0 & x \in ((mR + \Delta_2), +\infty] \end{cases} \quad (8)$$

It defines a zero mean variation of light intensity centered at I in a range of $-(mR + \Delta_2)$ to $(mR + \Delta_2)$ caused by eye-jittering. Fig.4 shows the intensity distribution vs neural codes.

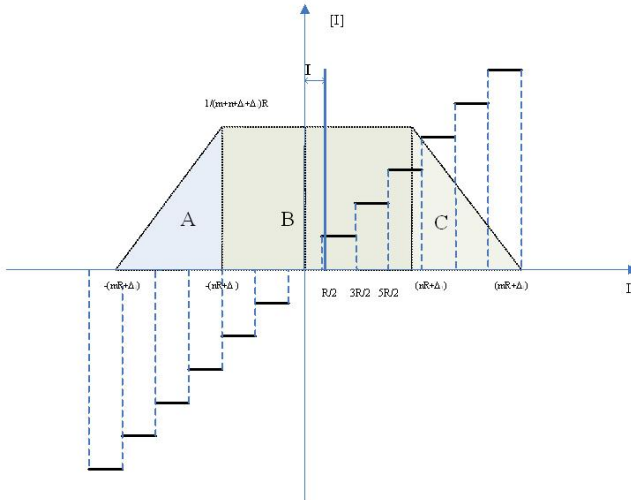


Fig.4. The probability density function vs. neural codes

And the corresponding distribution function can be obtained as

1) If $\Delta_1 + I > R/2$ and $\Delta_2 + I > R/2$, the bias of the first-order estimation in Fig.1 is

$$\begin{aligned} \Delta &= \frac{(2I - R)(\Delta_2 - \Delta_1)(R - \Delta_2 - \Delta_1)}{2[(m+n)R + (\Delta_2 + \Delta_1)][(m-n)R + (\Delta_2 - \Delta_1)]} \\ &= \frac{(2I - R)(\Delta_2 - \Delta_1)(R - \Delta_2 - \Delta_1)}{2[(a+b) + |a-b|][(a+b) + |a-b|]K^2} \end{aligned} \tag{9}$$

Similarly, we can have the following conclusions.

II) If $\Delta_2 + I \leq R/2$ and $\Delta_1 + I \geq R/2$,

$$\Delta = \frac{2(\Delta_2 - \Delta_1)(\Delta_2 + \Delta_1) + 4IR\left(\frac{R}{2} - \Delta_1\right) - R\left(\frac{R}{2} + I - \Delta_2\right)\left(\frac{R}{2} + I - \Delta_1\right)}{2[(a+b) + |a-b|][(a+b) + |a-b|]K^2} \tag{10}$$

III) if $\Delta_2 + A \geq R/2$ and $\Delta_1 + A \leq R/2$,

$$\Delta = \frac{-2I(\Delta_2 - \Delta_1)(\Delta_2 + \Delta_1) + R(\Delta_2 + I - \frac{1}{2}R)}{2[(a+b) + |a-b|][(a+b) + |a-b|]K^2} \tag{11}$$

IV) if $\Delta_2 + I \leq R/2$ and $\Delta_1 + I \leq R/2$,

$$\Delta = \frac{-(\Delta_2 - \Delta_1)(\Delta_2 + \Delta_1)R}{2[(a+b) + |a-b|][(a+b) + |a-b|]K^2} \tag{12}$$

It can be observed from I) to IV) that the statistic estimation of neural codes is a biased estimation. The bias Δ can be reduced by increasing K , i.e. the amplitude of eye movement. Less biased estimation can also be achieved if the perceived intensity has higher spatial intensity gradients, i.e. higher a and b . Therefore, we can control the bias to a desired level by adjusting eye movement if spatial gradients a and b are nonzero.

5 Simulations and Analysis

To illustrate the effect of the proposed visual neural network, we use a stationary image as visual stimuli to the photoreceptors and evaluate the first-order and the second order outputs of the neural networks respectively.

Here a fixed resolution frame, an 128*128 lattice, is used to scan a high resolution image, which simulates the digitalization of a continuous natural scene. Due to the limitation of the four-bits brightness resolution perception, the perceived image under-samples the light intensity. As in Fig.5(B), not only the color and contrast are deteriorated, but also spatial information is lost by biased wrong gray value. However, if

the random eye jitter is introduced, the color information and the details of the image could be refined by statistical window, as in Fig.5(C). Furthermore, increasing sampling times in statistical window will enhance the image quality as in Fig.5 (D), as the variation and bias in each pixel descend.



Fig. 5. Comparison on within and without fixational eye movement

Introducing eye-movement can further benefit image edge or contour extraction using the second-order neural output. Increasing amplitude of eye movement will generally make the variances around edges or contours more prominent. Fig.5(B). depicts how the amplitude of eye movement affects the second order output of the neural network, increasing amplitude leads to strengthened edge signals.

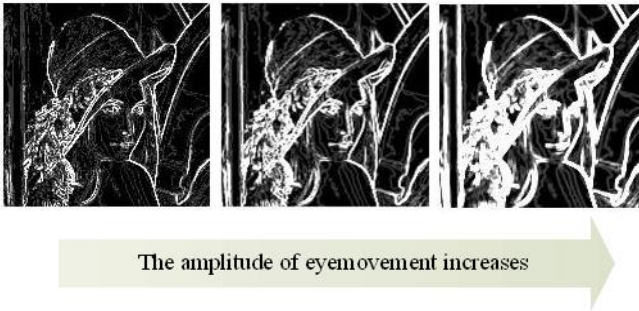


Fig. 6. The second-order statistics with increasing amplitude of eye movement

The capabilities of enhancement of fidelity in smooth area and prominence of high-frequency edges are the reason for introducing eye movement to enhance image acuity. However, in a dark environment or an area with uniform brightness, random jittering would not be able to enhance visual acuity due to the low contrast, i.e. lower a and b . In order to reduce the bias, a higher amplitude K is generally required. It coincides with the physiological findings reported in [14] that the fixational eye-movement is less frequent in the dark and the amplitude is larger if eye movement happens. However, a larger amplitude of fixational eye-movement does not always contribute to enhancement of visual acuity. There is a tradeoff between bias and variance of the estimation. While the vibration amplitude of the fixational eye-movement

increases, the noise of the neural outputs will boom as the analysis in section 4. From (7), it has to take longer time for statistics, i.e. increasing T , in order to compensate for the side effect of increasing noise. At the same time, the extracted contours from the second order outputs will be blurred as increasing the amplitude of eye-jittering. This trade-off can be depicted in Fig.7, increasing the random jittering will result in decreasing the estimation bias but increasing the variance. There is a need to select an optimum movement signal [10].

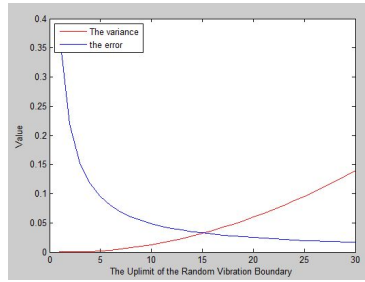


Fig.7. The trade-off between bias and variance

To solve this dilemma, we define an objective function to describe the tradeoff.

$$J_K = w_1 (\Delta)_K^2 + w_2 \left(\text{var}(\bar{X}) \right)_K^2 \tag{13}$$

where w_1 and w_2 are weights for bias and variance respectively; JK is determined by the exerted amplitude K and image gradients. Therefore, an optimal K can be expected for the tradeoff by solving the optimization. Thus we may further hypothesize that a visual system is able to adjust its eye movement pattern to adapt to external light stimuli for achieving a desired visual acuity.

6 Conclusions

It has been shown that the image acuity, both resolution and sharpness, can be improved by introducing eye movement when fixation. As eye movement generates additive random stimuli to photoreceptors, a visual neural model was presented to take into account temporal statistics of firing neurons. The statistical characteristics of the firing neurons under random eye movement are obtained. The visual neural model explained some physiological findings of eye-movements. It was derived from the visual neural model that eye movement requires a tradeoff between bias and variance for enhancement of image acuity, i.e. between expected fidelity and response time. It was formulated as a two-objective optimization problem and will be taken as a future research work to include an adaptive mechanism into the visual neural model. Simulations were carried out to demonstrate the capability of image acuity enhancement from eye movement using the proposed neural networks, which demonstrated the

application potential to develop high-acuity vision systems by introducing active and adaptive vibration.

References

1. Ditchburn, R.W., Ginsborg, B.L.: Involuntary Eye Movements During Fixation. *J. Physiol.* 119(1), 1–17 (1953)
2. Ginsborg, B.L., Maurice, D.M.: Involuntary Movements of the Eye During Fixation and Blinking. *Br. J. Ophthalmol.* 43(7), 435–437 (1959)
3. Spauschus, A., Marsden, J., Halliday, D.M., Rosenberg, J.R., Brown, P.: The Origin of Ocular Microtremor in Man. *Experimental Brain Research* 126(4), 556–562 (1999)
4. García-Pérez, M., Peli, E.: Motion Perception under Involuntary Eye Vibration. In: *European Conference on Visual Perception*, Glasgow, UK (2002)
5. Bolger, C., Bojanic, S., Sheahan, N.F., Coakley, D., Malone, J.F.: Dominant Frequency Content of Ocular Microtremor from Normal Subjects. *Vision Research* 39(11), 1911–1915 (1999)
6. Wang, L., Li, Y.J., Zhang, K.: Fixation Eye Movement Research and the Simulation of Retinal Information Processing Mechanism, <http://www.paper.edu.cn>
7. Landolt, O., Mitros, A.: Visual Sensor with Resolution Enhancement by Mechanical Vibrations. *Autonomous Robots* 11(3), 233–239 (2001)
8. Hennig, M.H., Wörgötter, F.: Eye Micro-movements Improve Stimulus Detection Beyond the Nyquist Limit in the Peripheral Retina. *Advances in Neural Information Processing Systems* 16 (2003)
9. Hennig, M.H., Kerscher, N.J., Funke, K., Wörgötter, F.: Stochastic Resonance in Visual Cortical Neurons: Does the Eye-tremor Actually Improve Visual Acuity? *Neurocomputing* 44, 115–120 (2002)
10. Greenwood, P.E., Lansky, P.: Optimum Signal in a Simple Neuronal Model with Signal-dependent Noise. *Biological Cybernetics* 92(3), 199–205 (2005)
11. Cherif, R., Nait-Ali, A., Motsch, J.F., Krebs, M.O.: A Parametric Analysis of Eye Tremor Movement during Ocular Fixation. In: *Proceedings of the 25th Annual International Conference of the IEEE on Applied to schizophrenia, Engineering in Medicine and Biology Society*, vol. 3, pp. 2710–2713 (2003)
12. Martinez-Conde, S., Macknik, S.L., Hubel, D.H.: The Role of Fixational Eye Movements in Visual Perception. *Nature Reviews Neuroscience* 5, 229–240 (2004)
13. Martinez-Conde, S., Macknik, S.L., Hubel, D.H.: Microsaccadic Eye Movements and Firing of Single Cells in the Striate Cortex of Macaque Monkeys. *Nature Neuroscience* 3, 251–258 (2000)
14. Greschner, M., Bingard, M., Rujan, P., Ammermuller, J.: Retinal Ganglion Cell Synchronization by Fixational Eye Movements Improves Feature Estimation. *Nature Neuroscience* 5, 341–347 (2002)

A General-Purpose FPGA-Based Reconfigurable Platform for Video and Image Processing

Jie Li¹, Haibo He¹, Hong Man¹, and Sachi Desai²

¹ Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030 USA
{jli8,hhe,Hong.Man}@stevens.edu

² U.S. Army, Armament Research, Development and Engineering Center (ARDEC), Picatinny, NJ 07806
{sachi.desai}@us.army.mil

Abstract. This paper presents a general-purpose, multi-task, and reconfigurable platform for video and image processing. With the increasing requirements of processing power in many of today's video and image processing applications, it is important to go beyond the software implementation to provide a real-time, low cost, high performance, and scalable hardware platform. In this paper, we propose a system by using the powerful parallel processing architecture in the Field Programmable Gate Array (FPGA) to achieve this objective. Based on the proposed system level architecture and design strategies, a prototype system is developed based on the Xilinx Virtex-II FPGA with the integration of embedded processor, memory control and interface technologies. Our system includes different functional modules, such as edge detection, zoom-in and zoom-out functions, which provides the flexibility of using this system as a general video processing platform according to different application requirements. The final system utilizes about 20% of logic resource, 50% of memory on chip, and has total power consumption around 203 mw.

Keywords: Reconfigurable system, FPGA design, video and image processing, edge detection, image scaling.

1 Introduction

Video and image processing plays a critical role in today's consumer electronics society. Over the past decades, we have witnessed a tremendous technology evolution on such technologies. With the continuous technological innovations in this area, many new opportunities as well as challenges have been raised in the consumer electronics research community. For instance, the switching of video technology from standard definition (SD) to high definition (HD) requires a six-time increase in data processing [1]. Video surveillance is also changing from the conventional common intermediate format (CIF) to the D1 standard [1]. To this end, the increased requirements of processing power, such as bandwidth, real-time computation, low latency, high throughput and low power consumption, have been the major focuses of the research efforts in the community from academia and industry as well.

Traditional computations of video and image processing normally require high-performance custom hardware implementations. Although such application specific integrated circuits (ASIC) can generally provide high density and power efficient systems, it requires a complicated design process. Fortunately, the developments of submicron and deep-submicron technology have enabled the FPGA to be a powerful hardware platform for many applications [2][3]. For instance, an image-scaling algorithm using an area pixel model, named Win-scale, has been implemented in FPGA [4]. This implementation has five functional blocks, including prescaler, line buffer, winfilter, filter window interpolator, and filter coefficients generator. The final system has a total of 29,000 NAND-equivalent gate counts after synthesis. In [5], a FPGA implementation of the SPIHT, a wavelet-based image compression coder, was presented. Various discrete wavelet transform architectures were implemented based on the WildStar processor board with three Xilinx Virtex 2000E FPGAs. It was reported that the final system achieved a 450-time speedup versus a microprocessor solution. A real-time optical-flow processing system based on FPGA was presented in [6]. This system used a pipelined architecture and was implemented in the Xilinx XCV2000E Virtex FPGA. Both software simulation and hardware testing results illustrated the effectiveness of this method. In [7], a video processing engine with accurate motion detection and sawtooth artifacts remove capabilities for LCD TV was proposed. This system used both temporal and spatial information of video fields for accurate motion detection, and adopted a window expanding search approach for low-angle edge detection. System level FPGA architecture and various emulation and verification results were presented. In [8], an approach for implementing a low cost TV set-top box (STB) capable of expanding sign images and decoding closed caption data was presented. The key FPGA implementation of this system includes a 27 MHz 8 bits microcontroller, an image expander and an on screen display unit. It was reported that this system can decode both Thai and English captions. In [9], a multi-window partial buffering (MWPB) scheme for 2-D convolvers for image and video processing was presented. Comparing to full buffering schemes, the proposed MWPB scheme has a good balance between on-chip resource utilization and external memory bus bandwidth. Other works of FPGA-based video and image processing systems include the motion-JPEG2000 for a digital cinema camera system [10], a fully multiplexed frequency-planar filter module (FMFPM) based on Xilinx Virtex-II FPGA [11], and others.

Most of the existing FPGA-based designs are focused on implementing specific algorithms for domain-specific applications. There are very few, if any, general-purpose hardware platforms to support and facilitate complex video and image processing. Therefore, a reconfigurable, low cost and scalable platform is highly desirable for the community. To this end, we propose a FPGA-based system including embedded processor, memory control and interface technologies to accomplish this objective.

The rest of this paper is organized as follows. Section II presents the system level architecture of the proposed FPGA platform. In section III, we present the

detailed implementation and design architectures of different functional modules. Section IV presents various experimental results under different video processing applications. Finally, a conclusion and a brief discussion on future research directions are discussed in section V.

2 FPGA-Based Video and Image Processing Platform

2.1 System Level Architecture

Generally speaking, a complex video application requires simultaneous data processing among different modules. The highly parallel data operation characteristic of FPGA provides a unique advantage of its application for such a purpose. According to different application requirements and specifications, different categories of FPGA chips can be used. In our current design, we use a low-cost high-end FPGA product, the Virtex-II Pro family (XC2VP30) as the prototype platform. Fabricated in 0.13 μ m process technology, the Virtex-II Pro family provides a good platform to meet different design requirements. For instance, the XC2VP30 FPGA includes dual Power-PC cores, over thirty thousand logic elements and 2Mbits embedded RAM [12]. Compared to the conventional DSP based design, the XC2VP30 FPGA can efficiently implement the multiply and accumulate (MAC) operations in parallel, and the behavior of each processor or peripheral core can be customized. Fig. 1 provides a system level architecture of the proposed video processing platform.

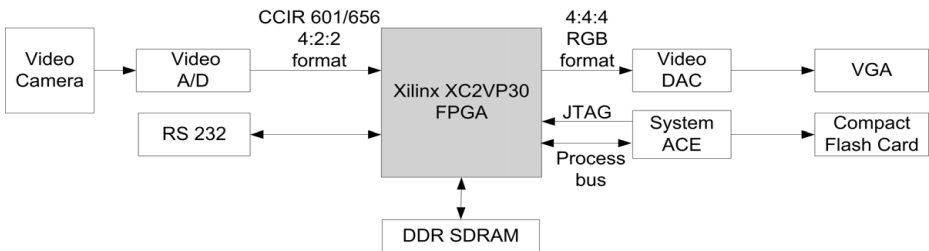


Fig. 1. The proposed system level architecture

In our system, a video analogue to digital conversion (ADC) board is used to capture the national television system committee (NTSC) signal and digitize it into CCIR 601/656 format. The architecture in Fig. 1 provides the flexibility of implementing different functional modules for video and image processing. In our current design, we have implemented three processing functions: zoom-in, zoom-out and edge-detection. Fig. 2 shows the data processing flow of the proposed system. One can easily extend this architecture to include more modules, or to test their own design concepts and algorithms based on this platform.

From Fig. 2 one can see, the FPGA implementation of the proposed system includes five major functional modules (the highlighted ellipses): the user-specific

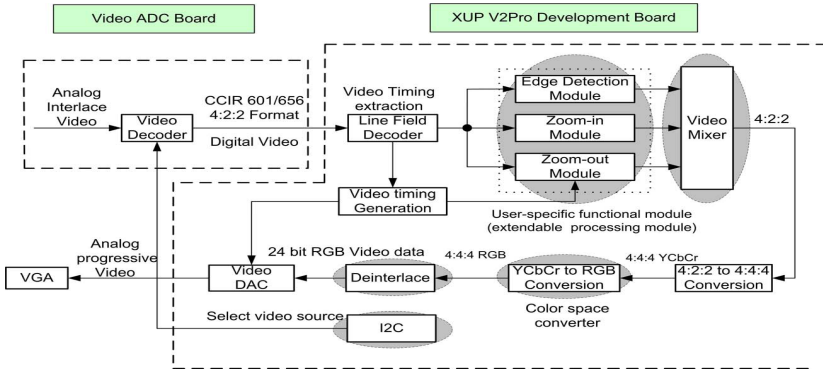


Fig. 2. Data processing flow of the proposed system

functional modules, the video mixer module, the color space converter module, the de-interlace module and the inter-integrated circuit (I2C) configuration module. The user-specific functional module implements most of the functionalities according to different video processing applications. This functional box can be extended in different application scenarios. The video mixer module can mix different video layers by the Alpha blending mixer function. This module supports both the picture-in-picture mixing and image blending. Each video layer can be independently displayed at running time. The color space converter module transforms the incoming video data between color spaces, which are specified by three coordinate values. This module supports the pre-defined conversions between standard color spaces, and allows user-specified coefficients to translate between any two three-valued color spaces. Interlaced video is commonly used in television standards such as phase alternation line (PAL) and NTSC. However, progressive video is required for LCD displays. Therefore, the de-interlace module converts interlaced video to progressive video. We use the embedded PowerPC405 microprocessor to achieve the I2C configuration function by programming the operational model of the analog device, the ADV7183B video decoder on the daughter card.

From Fig. 2 one can see, one of the advantages of the proposed system is that it provides an extendable module to implement different functionalities according to different application requirements. This provides the flexibility of using this system as a general-purpose video and image processing platform across different application domains. In our current research, we implement the edge detection and scaling (zoom-in and zoom-out) functions, which are important procedures in many complex video processing applications.

2.2 Four-Direction Edge Detection

Edge detection is a fundamental and critical technique in most image processing applications to obtain useful information before feature extraction and

object segmentation. This process detects outlines of an object and boundaries between objects and the background. In this research, we implement the four-direction Sobel operator [13] for edge detection. The detection resolutions and filter coefficients can be dynamically changed during the running time.

Generally speaking, the Sobel operator is based on a two-dimensional spatial gradient measurement on an image to detect the edges. This is implemented by calculating the convolutions of the image with a filter mask (convolution kernel) to calculate the approximate gradient magnitude [13]. Typically, the convolution kernel is moved pixel-by-pixel and line-by-line across the image, which can be defined as:

$$h [i, j] = f [i, j] * g [i, j] = \sum_{k=0}^{n-1} \sum_{l=0}^{m-1} f [k, l] g [i - k, j - l] \tag{1}$$

Where $g (i, j)$ represents the convolution kernel, n and m is the size of the convolution kernel at two dimensions, and $f (i, j)$ and $h (i, j)$ represents the original and filtered image, respectively.

A 3 by 3 kernel is used in our design to produce the map of intensity gradients. This is implemented by using the four-direction gradients calculated by convoluting the source video frame with the four-direction kernels. Fig. 3 illustrates this idea.

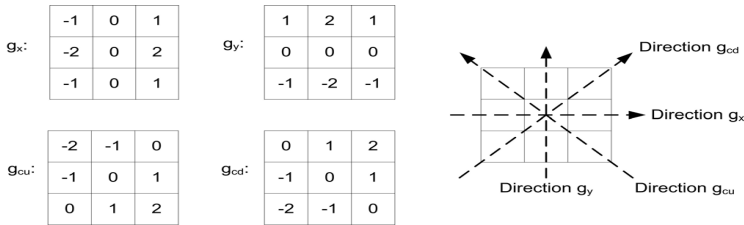


Fig. 3. Four-direction edge detection

In order to implement this four-direction edge detection, a generic 2-D image filter is proposed in Fig. 4. In this design, two line buffers and six registers are used to store the data flow and provide access to the neighborhood pixels. The incoming pixels are shifted through line buffers to create a delay line, which are sent to the filter array simultaneously with pixels from all the relevant video lines. At each filter node, the pixel is multiplied with the appropriate filter coefficients as indicated in Fig. 3. All the multiplier results are added together at the adder tree to produce the filter middle point output result. From Fig. 4 one can see, four additions and nine multiplications are needed to calculate the output value of the convolution.

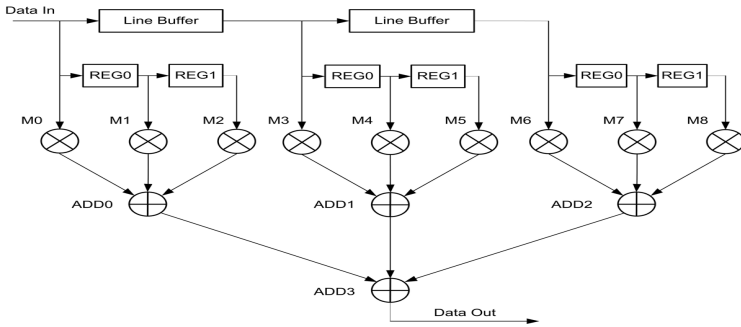


Fig. 4. Design of the four-direction edge detection

2.3 Scaling Functionalities: Image Zoom-In and Zoom-Out

Scaling is another widely used technique in many video processing applications. In this research, we implement the zoom-in and zoom-out functions in the extendable functional module.

As far as the zoom-in function is concerned, there are several popular algorithms such as nearest neighbor method and bilinear interpolation method [14]. Our current design supports both methods and can be configured to change resolutions and/or filter coefficients at running time. As an example, Fig. 5 gives a detailed design architecture of the bilinear interpolation method. Without loss of generality, we assume the upscale factor is two and one need to zoom in as four times as the original image. Fig. 5 illustrates the method that is used to generate new pixels and new lines of the image. First, new pixels between line n and line $n + 1$ are generated with a combination factor of $1/2$. Then, new pixels between the two vertical pixel lines are created. In our design, two video frame buffers are used: one is used to store the luminance signals and the other one is used to store the chroma signals.

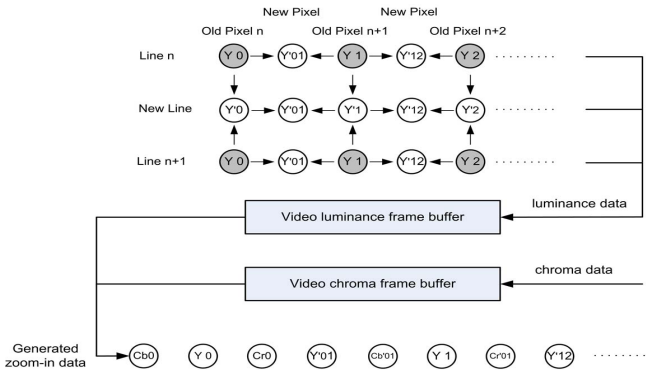


Fig. 5. Design of the zoom-in function for video processing

Fig. 6 illustrates the idea of implementing the zoom-out function. In order to eliminate the frequency mixing effect, the incoming images are first passed through a low-pass filter. The new pixels are then calculated by bilinear interpolation method. Assuming the zoom-out image is a quarter of the original image, Fig. 6 illustrates the data flow to implement this, where C_b and C_r represent video chroma data, and Y represents video luminance data.

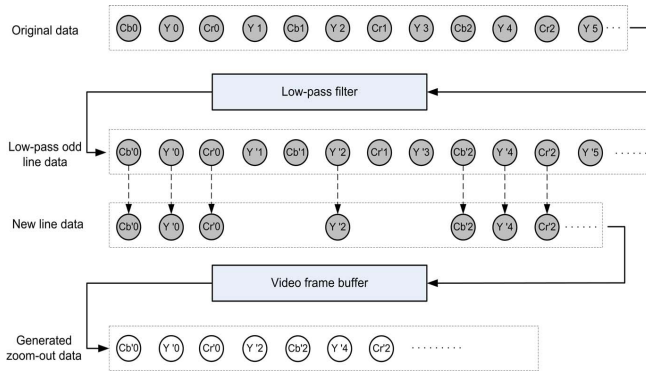


Fig. 6. Design of the Zoom-out function for video images

3 System Implementation and Experimental Results

3.1 System Implementation

We implement the entire platform based on the Xilinx Virtex-II Pro development system. Fig. 7 shows the hardware platform with major components. The on board XC2VP30 FPGA chip has about 30,816 logic cells, 136 18-bit multipliers, 2,448Kb of block RAM, and two PowerPC Processors. The DDR SDRAM DIMM can support up to 2Gbytes of RAM. This board also has many useful interface ports, such as the 10/100 Ethernet port, compact flash card slot, XSGA video port, RS-232 port, and others. It also has various expansion connectors to expand the usability of this board to meet the requirements of different video and image processing applications. Our major purpose of this system is to implement the entire hardware platform to provide a general solution for video and image processing, and demonstrate its effectiveness through various application scenarios.

To verify the timing and logic functions, the entire system is simulated by the Xilinx Integrated Software Environment (ISE 9.1i) toolsets for extensive simulation and logic analysis. Fig. 8 shows a snapshot of the system logic and timing simulation results. The system operation clock is 27MHz (the clk_27 signal). The pcount signal counts the number of line pixels, and the firstline_data, secondline_data and thirdline_data represent the input video data of three lines. We operate the line buffer through fifo_wen and fifo_ren signal, which generates the

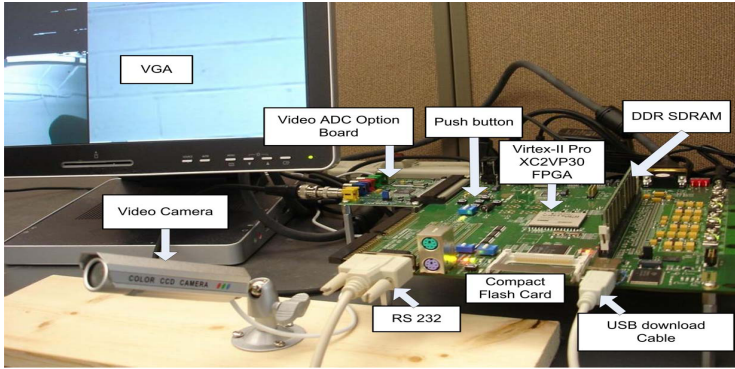


Fig. 7. The proposed FPGA platform

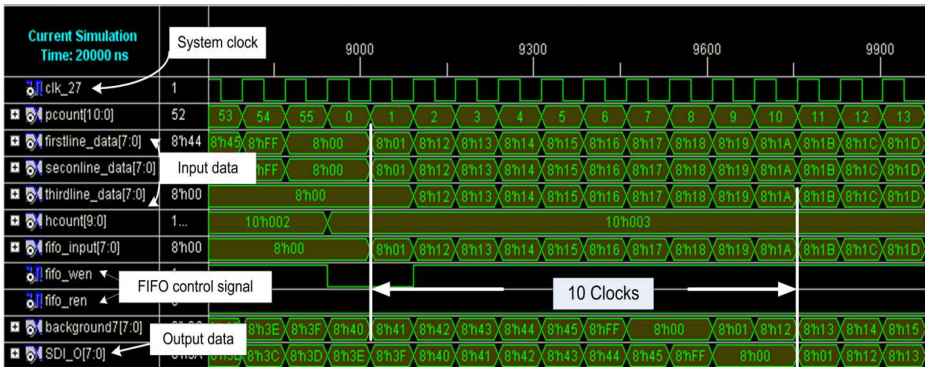


Fig. 8. A snapshot of the system logical simulation

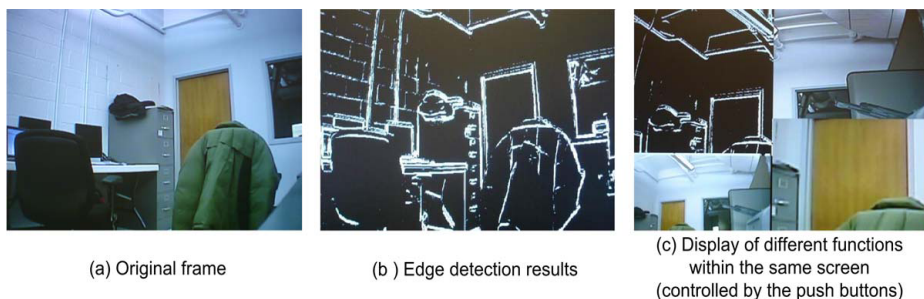
background signal (background7) by delaying proper number of clocks from the original input video stream. By mixing the background signal and the processed video data (the f_data signal), we can get the final output signal (the SDI_O signal). From Fig. 8 one can see that a total of 10 clocks processing time is needed for one pixel operation. Table 1 summarizes the major resource utilization characteristics of the final system, from which one can see the final system utilizes about 20% of logic resource, 50% of memory on chip, and has total power consumption around 203mw.

3.2 Experimental Results

In this section, we demonstrate the effectiveness of the hardware system for different video processing applications. The input video can either be obtained through a camera system or other video devices. Fig. 9 shows the results of using the camera system to capture image data from different environments. Fig. 9 (a) shows the original image and Fig. 9 (b) illustrates the effects of the

Table 1. Resource utilization of the entire system

Hardware resource	Available	Used	Utilization
Number of occupied Slices	13696	2869	20%
Total Number of 4 input LUTs	27392	5293	19%
Number of bonded IOBs	556	42	7%
Number of PPC405s	2	1	50%
Number of Block RAMs	136	70	51%
Number of MULT18X18s	136	5	3%
Number of GCLKs	16	2	12%

**Fig. 9.** System performance based on the camera input data

edge detection function. Fig. 9 (c) demonstrates all the functional modules in the same window, including the edge detection, zoom-in and zoom-out. All these functions can be controlled easily by the push buttons on the FPGA board (see Fig. 7 for system details).

4 Conclusions and Future Work

In this paper, we propose a FPGA-based prototype system for general purpose and multi-task video and image processing. System level hardware architecture and detailed design strategies are presented. The final system is implemented using the Xilinx Virtex-II Pro development system with an onboard XC2VP30 FPGA chip. Synthesized results indicate the overall system utilizes only about 20% of logic resource, 50% of memory on chip, and has total power consumption around 203 mw. This system provides a scalable and real-time reconfigurable platform to meet the requirements for many video processing applications. Furthermore, the reconfigurable and extendable characteristics of this system allow it to be easily modified to embed into different video and image processing scenarios. The effectiveness of the proposed prototype has been demonstrated by various experimental results.

In the future work, it would be interesting to integrate more complicated video processing modules into this platform. For instance, based on the edge

detection function implemented in this research, it will be useful to implement a robust objects recognition algorithm into this system. In addition, since machine learning techniques have been extensively used for video and image processing, it would be interesting to develop various learning algorithms based on this prototype. For instance, we are currently designing a FPGA-based incremental learning system for video applications. The key idea is to develop an incremental learning architecture in hardware to learn and accumulate knowledge for multiple objects recognition and localization. Motivated by our research in this paper, we believe that such a FPGA-based system will provide a power platform for many real-world video and image processing applications.

Acknowledgments. This work was supported in part by the US Army Armanent Research, Development and Engineering Center (ARDEC), Picatinny, under Grant No. W15QKN-05-D-0011.

References

1. Video and Image Processing Design Using FPGAs, Altera Corporation, <http://www.altera.com/literature/wp/wp-video0306.pdf>
2. Benkrid, K., Crookes, D., Benkrid, A.: Towards a General Framework for FPGA Based Image Processing Using Hardware Skeletons. *Parallel Computing* 28(7-8), 1141–1154 (2002)
3. Johnston, C.T., Gribbon, K.T., Bailey, D.G.: Implementing Image Processing Algorithms on FPGAs. In: *Proc. Electronics New Zealand Conference*, Palmerston North, New Zealand, pp. 118–123 (2004)
4. Kim, C.H., Seong, S.M., Lee, J.A., Kim, L.S.: An Image-Scaling Algorithm Using an Area Pixel Model. *IEEE Trans. Circuits and Systems for Video Technology* 13(6), 549–553 (2003)
5. Fry, T.W., Hauck, S.A.: SPIHT Image Compression on FPGAs. *IEEE Trans. Circuits and Systems for Video Technology* 15(9), 1138–1147 (2005)
6. Daz, J., Ros, E., Pelayo, F., Ortigosa, E.M., Mota, S.: FPGA-Based Real-Time Optical-Flow System. *IEEE Trans. Circuits and Systems for Video Technology* 16(2), 274–279 (2006)
7. Ku, C.C., Liang, R.K.: Accurate Motion Detection and Sawtooth Artifacts Remove Video Processing Engine for LCD TV. *IEEE Trans. Consumer Electronics* 50(4), 1194–1201 (2004)
8. Leelarasmee, E.: A TV Sign Image Expander with Built-in Closed Caption Decoder. *IEEE Trans. Consumer Electronics* 51(2), 682–687 (2005)
9. Zhang, H., Xia, M., Hu, G.: A Multiwindow Partial Buffering Scheme for FPGA-Based 2-D Convolvers. *IEEE Trans. Circuits and Systems, Part II* 54(2), 200–204 (2007)
10. Fel, S., Fttinger, G., Mohr, J.: Motion JPEG2000 for High Quality Video Systems. *IEEE Trans. Consumer Electronics* 49(4), 787–791 (2003)
11. Madanayake, A., Bruton, L.: A Fully Multiplexed First-Order Frequency-Planar Module for Fan, Beam, and Cone Plane-Wave Filters. *IEEE Trans. Circuits and Systems, Part II* 53(8), 697–701 (2006)

12. Virtex-II Pro family (XC2VP30), Data sheet,
http://www.xilinx.com/products/silicon_solutions/fpgas/virtex/virtex_ii_pro_fpgas/index.htm
13. Dhawan, A.P.: Medical Image Analysis, pp. 175–210. Wiley-interscience Press, Hoboken (2003)
14. Jahne, B.: Digital Image Processing, 5th edn., pp. 427–440. Springer, Berlin (2002)

Image Analysis by Modified Krawtchouk Moments

Luo Zhu¹, Jiaping Liao¹, Xiaoqin Tong¹, Li Luo¹, Bo Fu¹, Guojun Zhang²

¹ School of Electrical and Electronic Engineering,
Hubei University of Technology, Wuhan 430068, China

² School of Mechanical Science and Engineering,
Huazhong University of Science and Technology, Wuhan 430074, China
luozhu812@yahoo.com.cn, jpliao@mail.hbut.edu.cn,
tong85214@163.com, luoli5920@163.com, fubofanxx@yahoo.com.cn,
zgj@mail.hust.edu.cn

Abstract. In the paper, a set of modified Krawtchouk moments with high accuracy and computational speed is introduced. Three computational aspects of Krawtchouk moments, which are weighted and normalized Krawtchouk polynomials, symmetry and recurrence relation, are discussed respectively. Firstly, by normalizing the Krawtchouk polynomials with the weight functions and norms, the values of the polynomials are limited to a smaller range than those of the classical polynomials. Secondly, three symmetrical properties are used to simplify the computational complexities of the high-order moments by reducing the modified polynomials by a factor of eight and lower the highest order of the calculated polynomials from N to $N/2-1$. Thirdly, the classical recursive relations are modified to calculate the normalized polynomials when the order N goes larger. Finally, the paper demonstrates the effectiveness of the proposed moments by using the method of image reconstruction.

Keywords: Krawtchouk moments, Krawtchouk polynomials, Image reconstruction, Symmetry, Recurrence relations.

1 Introduction

Moment functions have been used as feature characteristics in many fields of image processing [1-5]. In 1961, Hu introduced a set of moment invariants based on the theory of algebraic invariants, which are translation, scale and rotation independent. However, regular moments are not orthogonal and as a result, reconstructing the image from the moments is a difficult work indeed. Teague firstly introduced moments with orthogonal basis functions, with the additional property of minimal information redundancy in a moment set. In this class, Legendre and Zernike moments have been widely studied in the recent past [2-5].

Since the Zernike and Legendre polynomials are defined only inside the unit region, the calculation of those moments requires a coordinate transformation and proper approximation of the continuous moment integrals. Discrete orthogonal moments, such as Tchebichef moments [6-7], Hahn moments [8], Krawtchouk moments [9], are directly defined in the image coordinate space and retained the property of orthogonality in a moment set, which are hence expected to perform better than continuous moments.

However, when the moment order becomes large, the squared norm of the scale discrete orthogonal polynomials presents very small values, leading to numerical instabilities in the computed moments. Another problem encountered in the computation of discrete orthogonal moments of larger order is the propagation of numerical errors while using the recurrence relation to evaluate the polynomial values. Mukundan [7] and Liang [8] respectively did an analysis of numerical instabilities of Tchebichef moments and Hahn moments, and proposed the recurrence relations with respect to x axis to reduce the accumulation errors in the computation of high order polynomial values. Yap proposed Krawtchouk moments [9], which have the capability of being able to extract local features from any region-of-interest in an image. And when the moment order becomes larger, Krawtchouk moments present the same problems as the above two moments, which can not be solved by using the same method. Therefore, a new solution is proposed in the paper.

2 Krawtchouk Polynomials and Moments

2.1 Krawtchouk Polynomials

The definition of the n -th order classical Krawtchouk polynomial is expressed as

$$K_n(x; p, N) = \sum_{k=0}^n a_{k,n,p} x^k = {}_2F_1\left(-n, -x; -N; \frac{1}{p}\right), \tag{1}$$

where $x, n = 0, 1, 2, \dots, N, N > 0, p \in (0, 1), {}_2F_1$ is the hypergeometric function and defined as

$${}_2F_1(a, b; c; z) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{z^k}{k!}, \tag{2}$$

and $(a)_k$ is the Pochhammer symbol given by

$$(a)_k = a(a+1)(1+2)\dots(a+k-1) = \frac{\Gamma(a+k)}{\Gamma(a)}. \tag{3}$$

The set of $(N + 1)$ Krawtchouk polynomials $\{K_n(x; p, N)\}$ forms a complete set of discrete basis functions with weight function

$$w(x; p, N) = \binom{N}{x} p^x (1-p)^{N-x}, \tag{4}$$

and satisfies the orthogonality condition

$$\sum_{x=0}^N w(x; p, N) K_n(x; p, N) K_m(x; p, N) = \rho(n; p, N) \delta_{nm}, \tag{5}$$

where $n, m = 1, 2, \dots, N, \delta_{nm}$ is the Kronecher function, i.e., and

$$\rho(n; p, N) = (-1)^n \left(\frac{1-p}{p} \right)^n \frac{n!}{(-N)_n}. \tag{6}$$

In order to make the computation of the polynomials less demanding on the processor, the recurrence relation can be used to avoid overflowing for mathematical functions like the hypergeometric function and gamma functions. The conventional three-term recursion of the Krawtchouk polynomials $K_n(x; p, N)$ with respect to n is

$$p(N-n+1)K_n(x, p, N) = (Np-2np+2p+n-1-x)K_{n-1}(x, p, N) - (n-1)(1-p)K_{n-2}(x, p, N) \tag{7}$$

with $K_0(x; p, N) = 1$ and $K_1(x; p, N) = 1 - \frac{x}{Np}$.

2.2 Krawtchouk Moments

Krawtchouk moments have the interesting property of being able to extract local features of an image. The Krawtchouk moments of order $(n + m)$ in terms of Krawtchouk polynomials, for an image with intensity function $f(x, y)$, is defined as

$$Q_{nm} = \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} K_n(x; p_1, N) \sqrt{w(x; p_1, N)} K_m(y; p_2, M) \sqrt{w(y; p_2, M)} f(x, y). \tag{8}$$

The parameters N and M are substituted with $N - 1$ and $M - 1$ respectively to match the $N \times M$ pixel points of an image. And the image density $\hat{f}(x, y)$ reconstructed by the moments can be expressed as

$$\hat{f}(x, y) = \frac{\sum_{x=0}^{N-1} \sum_{y=0}^{M-1} Q_{nm} K_n(x; p_1, N) \sqrt{w(x; p_1, N)} K_m(y; p_2, M) \sqrt{w(y; p_2, M)}}{\rho(n; p_1, N) \rho(m; p_2, M)}. \tag{9}$$

In order to show the image representation capability of Krawtchouk moments, an objective measure is used to characterize the error between the original image $f(x, y)$ and the reconstructed image $\hat{f}(x, y)$ is defined as follows

$$\varepsilon = \sqrt{\sum_{x=0}^{N-1} \sum_{y=0}^{M-1} [f(x, y) - \hat{f}(x, y)]^2}. \tag{10}$$

3 Computations Aspects of Krawtchouk Moments

As the moment order becomes larger, Krawtchouk moments will present numerical instabilities like Tchebichef moments [7] and Hahn moments [8], which can not be solved by using the recurrence relation with respect to x . For Krawtchouk moments, the recursion on x axis has the same style as that on n axis. And so, some theoretical framework will be discussed in this section.

3.1 The Modified Krawtchouk Polynomials

The conventional method of avoiding numerical fluctuations for moment computations is by means of normalization by the norm. However, both the values of the weight $w(x; p, N)$ and Krawtchouk polynomials have to be calculated for Krawtchouk moments. When the order N goes larger, the values of the weight will tend to zero as limit and the polynomial values increase sharply. The extreme values lead to numerical instability due to the length limitation of the storage in the computer. If the weight is used as a scale factor to counteract the increment of the polynomial value for a large value of n , the numerical fluctuations will be restrained in some extent. Therefore to achieve the numerical stability, a new set of weighted and normalized Krawtchouk polynomials $\{\bar{K}_n(x; p, N)\}$ with respect to the norm is defined as

$$\bar{K}_n(x; p, N) = K_n(x; p, N) \sqrt{\frac{w(x; p, N)}{\rho(n; p, N)}}. \tag{11}$$

Then the Krawtchouk moments of order $(n + m)$ in terms of weighted and normalized Krawtchouk polynomials, for an $N \times N$ image with intensity function $f(x, y)$, can be obtained easily as follows

$$Q_{nm} = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \bar{K}_n(x; p_1, N) \bar{K}_m(y; p_2, N) f(x, y). \tag{12}$$

And the reconstructed image $\hat{f}(x, y)$ is expressed as

$$\hat{f}(x, y) = \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} Q_{nm} \bar{K}_n(x; p_1, N) \bar{K}_m(y; p_2, N). \tag{13}$$

3.2 Symmetry

Both the computation time and the computation error of the Krawtchouk moments can be reduced considerably by using the symmetry property to reduce computational complexities of the high-order moments. In [9], Yap has just introduced the symmetry with respect to x . However, the proposed polynomials have the symmetry properties not only on x orientation but also on the order n and the diagonal $n=x$. The three symmetry properties of the weighted and normalized Krawtchouk polynomials can be easily derived from the equations of the classical polynomials and listed as follows:

(a) The symmetry relation on x axis introduced by the following style

$$\bar{K}_n(x, p, N) = (-1)^n \bar{K}_n(N - x, p, N). \tag{14}$$

(b) The symmetry equation on n axis is expressed as

$$\bar{K}_n(x, p, N) = (-1)^x \bar{K}_{N-n}(x, p, N). \tag{15}$$

(c) The third symmetric property on diagonal is expressed as follows

$$\bar{K}_n(x, p, N) = \bar{K}_x(n, p, N). \tag{16}$$

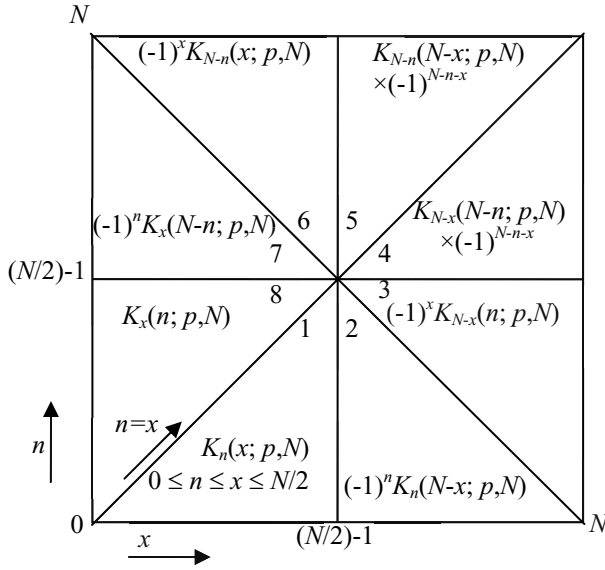


Fig. 1. Symmetrical properties of Krawtchouk polynomials

The above three symmetric properties suggests the subdivision of the domain of the polynomials set $\{\bar{K}_n(x; p, N)\}$ (where N is even) into eight equal parts (Fig. 1), and performing the computation of the modified polynomials only in the first part where $0 \leq x, n \leq N/2$ and $n \leq x$ by using (14), (15), (16). The rest of the polynomials can be determined by using the above three symmetry properties. For a more detailed explanation, refer to [6] and [9].

Furthermore, the symmetry properties are also useful in lowering the highest calculated order from N to $N/2 - 1$, which would reduce the accumulation of the numerical errors largely. As an example, the expression for Krawtchouk moments in (12) can be modified by using the equation (14) and (15) as

$$\begin{aligned}
 Q_m &= \sum_{x=0}^{(N/2)-1} \sum_{y=0}^{(N/2)-1} \bar{K}_n(x; p, N) \bar{K}_m(y; p, N) \times \{f(x, y) + (-1)^n f(N-x, y) + (-1)^m f(x, N-y) + (-1)^{nm} f(N-x, N-y)\}, \\
 & \hspace{25em} \text{if } 0 \leq n, m < (N/2), \\
 &= \sum_{x=0}^{(N/2)-1} \sum_{y=0}^{(N/2)-1} (-1)^y \bar{K}_{N-n}(x; p, N) \bar{K}_m(y; p, N) \times \{f(x, y) + (-1)^n f(N-x, y) + (-1)^m f(x, N-y) + (-1)^{nm} f(N-x, N-y)\}, \\
 & \hspace{25em} \text{if } 0 \leq m < (N/2) \leq n, \\
 &= \sum_{x=0}^{(N/2)-1} \sum_{y=0}^{(N/2)-1} (-1)^y \bar{K}_n(x; p, N) \bar{K}_{N-m}(y; p, N) \times \{f(x, y) + (-1)^n f(N-x, y) + (-1)^m f(x, N-y) + (-1)^{nm} f(N-x, N-y)\}, \\
 & \hspace{25em} \text{if } 0 \leq n < (N/2) \leq m, \\
 &= \sum_{x=0}^{(N/2)-1} \sum_{y=0}^{(N/2)-1} (-1)^{x+y} \bar{K}_{N-n}(x; p, N) \bar{K}_{N-m}(y; p, N) \times \{f(x, y) + (-1)^n f(N-x, y) + (-1)^m f(x, N-y) + (-1)^{nm} f(N-x, N-y)\}, \\
 & \hspace{25em} \text{if } n, m \geq (N/2).
 \end{aligned} \tag{17}$$

3.3 Recurrence Relation

Based on the discussion in Section 3.1, we know that it is inadequate to ensure numerical stability by using the recurrence relation given in (7) to evaluate the weighted and normalized Krawtchouk polynomials. A new set of recursive relations to calculate the polynomials is proposed as follows

$$p(N-x+1)\bar{K}_n(x, p, N) = A(Np-2xp+2p+x-1-n)\bar{K}_n(x-1; p, N) - B(x-1)(1-p)\bar{K}_n(x-2; p, N) \quad (18)$$

where $A = \sqrt{\frac{(N-x+1)p}{x(1-p)}}$ and $B = \frac{p}{1-p} \sqrt{\frac{(N-x+1)(N-x+2)}{x(x-1)}}$,

with $\bar{K}_n(0; p, N) = \sqrt{\frac{p^n(1-p)^{N-n}N!}{n!(N-n!)}}$ and $\bar{K}_n(1; p, N) = (Np-n) \sqrt{\frac{1}{Np(1-p)}} \bar{K}_n(0; p, N)$.

4 Experiment and Analysis

The gray-level image in Fig. 2 are used to illustrate the problems associated with large-order classic Krawtchouk moments in image reconstruction, and also use to validate theoretical framework introduced in Section 3.

The reconstructions of the original image are recorded in Fig.3. Fig.3(a) is reconstructed images from the classical Krawtchouk moments and Fig.3(b) is obtained from our moments. Comparing the plots of the reconstruction error, as computed from (13), shown in Fig.4, the reconstruction errors of modified Krawtchouk moments are distinctly fewer than those of classical Krawtchouk moments as the order $N > 180$. It verifies that Krawtchouk moments have superior performance to the classical Krawtchouk moments when N goes larger in analyzing the large gray-level images accurately.



Fig. 2. A gray-level image of size 200×200 pixels used for reconstruction

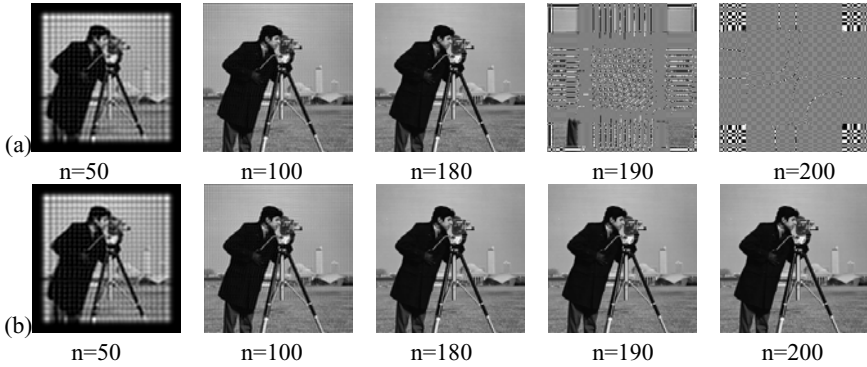


Fig. 3. Reconstructions with two moments (n is the maximum reconstructed order) (a) with conventional method; (b) with the proposed method

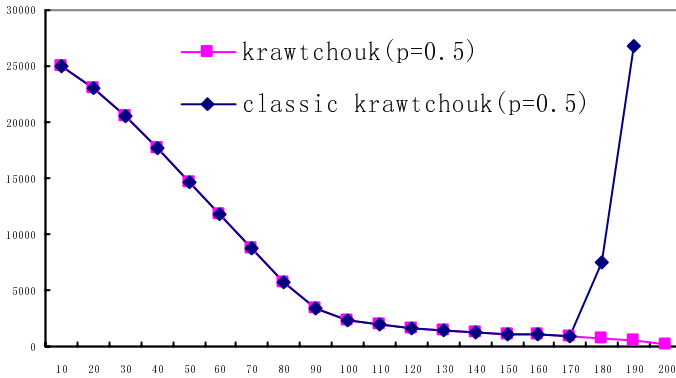


Fig. 4. Reconstruction errors for the gray-level image with four moments

5 Conclusion

This paper introduces a set of modified Krawtchouk moments with high accuracy and efficiency, which can be availably used as shape characteristics in the analysis of large images. In order to calculate the high order moments accurately, we respectively do an analysis of numerical instabilities of classical Krawtchouk moments in three computational aspects involving normalization, symmetry and recursion. Above all, the Krawtchouk polynomials are normalized by the weight functions and norms so that the values of the polynomials are limited to a smaller range than that of the classical polynomials. Additionally, three symmetrical properties of our polynomials are derived and used to reduce the computation of the polynomials by a factor of eight, by which both the computation time and the computation error of the Krawtchouk moments can be decreased considerably. And then, a modified recursive relation is correspondingly presented to evaluate the proposed polynomials when the order N goes larger. At Last,

the image reconstruction ability of our moments is compared with that of the classical Krawtchouk moments and the experiment results conclusively demonstrate the effectiveness of the proposed method as feature descriptors. Future work in the field of Krawtchouk moments is directed toward the identification of invariants, and feasibility studies on the use of Krawtchouk polynomials in two variables as basis functions.

Acknowledgment. This investigation is supported by the Project of National Natural Science Foundation of China (No. 60702079).

References

1. Dudani, S., Breeding, K., McGhee, R.: Aircraft Identification by Moment Invariants. *IEEE Trans. Comput.* 26, 39–45 (1977)
2. Mukundan, R., Ramakrishnan, K.R.: Fast Computation of Legendre and Zernike Moments. *Pattern Recognition* 28, 1433–1442 (1995)
3. Fu, B., Zhou, J.Z., Li, Y.H.: Image Analysis by Modified Legendre Moments. *Pattern Recognition* 40, 691–704 (2007)
4. Kamila, N.K., Mahapatra, S., Nanda, S.: Invariance Image Analysis Using Modified Zernike Moments. *Pattern Recognition Lett.* 26, 747–753 (2005)
5. Liao, S.X., Pawlak, M.: On the Accuracy of Zernike Moments for Image Analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 20, 254–266 (1998)
6. Mukundan, R., Ong, S.H., Lee, P.A.: Image Analysis by Tchebichef Moments. *IEEE Trans. Image Process.* 10, 1357–1364 (2001)
7. Mukundan, R.: Some Computational Aspects of Discrete Orthonormal Moments. *IEEE Trans. Image Process* 13, 1055–1059 (2004)
8. Liang, J., Shu, H.Z., Zhu, H.Q., Luo, L.M.: The Image by Discrete Orthogonal Moments. *Journal of Shanghai Jiaotong University* 40, 796–800 (2006)
9. Yap, P.T., Paramesra, R., Ong, S.H.: Image Analysis by Krawtchouk Moments. *IEEE Trans. Image Process* 12, 1367–1377 (2003)

Efficient Provable Secure ID-Based Directed Signature Scheme without Random Oracle

Jianhong Zhang^{1,2}, Yixian Yang², and Xinxin Niu²

¹ College of sciences, North China University of Technology, Beijing 100041, China

² School of information engineering, Beijing University of posts and telecommunications, Beijing 100876, China
jhzhangs@gmail.com, yxyang@bupt.edu.cn

Abstract. As a special signature, a directed signature is a type of signature with verification ability which is restricted. In a directed signature scheme, a designated verifier can exclusively verify the validity of a signature. If necessary, the designated verifier or the signer can prove the correction of a signature to a third party. Directed signature schemes are suitable for applications such as bill of tax and bill of health. In this paper, an ID-based directed signature scheme without random oracle is proposed by combining ID-based cryptology with Waters signature. We also give the syntax and security notion of ID-based directed signature without random oracle: unforgeability and invisibility. Finally, we show that the proposed scheme is unforgeable under the computational Diffie-Hellman assumption, and invisible under the Decisional Bilinear Diffie-Hellman assumption.

Keywords: Directed signature, ID-based, CDH problem, DBDH problem, Standard model.

1 Introduction

In ordinary digital signature schemes, anyone can verify the validity of a signature with signer's public key. However, in some scenarios, it is not necessary for anyone to be convinced a validity of signer's confidential message, since the signed message may contain a confidential agreement or a private information between the signer and the recipient. For example, signatures on medical records, tax information and most business transaction. To address this problem above, Chaum and Van Antwerpen introduced undeniable signature [1] which allowed a signer to have complete control over his signatures. Because undeniable signatures have various applications in the security of e-commerce, such as licensing software, auctions and electronic voting, many variants of undeniable signature appear, such as FDH undeniable signature [1] and threshold undeniable signature [2,3]. However, undeniable signatures are only verified with the cooperation of the signer. Thus, it is very inconvenient and impractical in real life. As an alternative approach to undeniable signatures, designated confirmer signature [4] was proposed by Chaum in 1994. In the scheme, a designated confirmer

signature allows certain designated parties to confirm the authenticity of a document without the need for the signer's input. At the same time, many signature types with controlled verifiability are proposed, such as limited verifier signature, designated verifier signature [5]. These schemes mainly focus on the ability of verification which is limited. However, we may meet the following situation:

A hospital A has issued a hospital record to the patient, Bob, in the form of hospital A 's digital signature. Bob then wants to exclusively verify these signatures with others knowing nothing about his state of illness. Otherwise, his state of illness is exposed. After a period time, he also needs to prove validity of his hospital record to other hospitals for cure. At the same time, hospital A also shares the ability and responsibility to acknowledge this hospital records when Bob may not be convenient to do so.

The aforementioned signature schemes with verifiability restriction seem to not be suitable for the above situation, as the verifier cannot prove validity of a signature to the others in a designated verifier signature and only the recipient can acknowledge a signature to a third party in a limited verifier signature. In [6], to solve the above problem, Lim and Lee proposed a new type of signature : directed signature, based on Guillou-Quisquater signature scheme. In 2005, Laguillaumie *et al* [7] studied the universally convertible directed signatures and gave a concrete scheme which was proven to be secure in the random oracle model [8]. In 2004, Sunder Lal *et.al* proposed a (t, n) threshold directed signature scheme [9] based on Shamir's threshold signature [10] and Schnorr's signature scheme [11]. Subsequently, Lu *et.al* also proposed a (t, n) threshold directed signature [3] based on elliptic curve, and they give a formal model of threshold directed signature. Finally, they gave a concrete scheme and showed that the scheme was existentially unforgeable in the random oracle model.

To the best of my knowledge, all the directed signature schemes are based on Public Key Infrastructure setting, and their securities are only proven secure in the random oracle model. It is well-known that security in the random oracle models does not imply security in the real world. Identity-based (ID-based) cryptography which was introduced by Shamir [12], has rapidly emerged in recent years and been widely applied . The prominent property of ID-based cryptography is that a user's public key can be any binary string which can identify the user, such as an email address. The idea was to eliminate public key certificates by using a public key that is bound to a users identity like an identity (ID) string of the entity involved (e.g. email address, telephone number, etc.). ID-based cryptography is supposed to provide a more convenient alternative to conventional public key infrastructure. Thus, many ID-based schemes [13,14] were proposed.

To solve the existing problems in the present directed signature schemes, an ID-based directed signature without random oracle is proposed by combining ID-based cryptology and directed signature in the paper. The idea of our scheme is based on that of Waters' signature scheme [15]. And we prove the scheme is secure against existential unforgeability attack based on the intractability of the CDH problem . We proceed to show that the scheme is secure against invisibility attack

based on the intractability of the decisional Bilinear Diffie-Hellman (DBDH) assumption. It is well-known that Waters’ signature is malleable, any one can produce a new signature according to a given signature δ on message m . Though our scheme is based on Waters’ signature, our scheme avoids the malleability of signature by including a hash function.

The rest of the paper is organized as follows: Section 2 briefly describes the necessary background concepts; Section 3 presents security model of ID-based directed signature scheme without random oracle; Our ID-based directed signature is proposed in section 4; Security proof and efficiency analysis of the scheme are given in section 5. Finally, we conclude our work.

2 Preliminaries

In this section, we first review some background on groups with efficiently computable bilinear pairing [16,17]. Then, we give the corresponding difficult mathematics problems which our scheme is based on.

Let \mathbb{G}_1 be a cyclic additive group generated by the generator P , whose order is a prime q , and \mathbb{G}_2 be a cyclic multiplicative group of the same prime order q . We assume that the discrete logarithm problems (DLP) in both \mathbb{G}_1 and \mathbb{G}_2 are hard. An admissible bilinear pairing e is defined as $e : \mathbb{G}_1 \times \mathbb{G}_1 \rightarrow \mathbb{G}_2$ with the following three properties:

- Bilinearity: If $P, Q \in \mathbb{G}_1$ and $a, b \in \mathbb{Z}_q^*$, then $e(aP, bQ) = e(P, Q)^{ab}$;
- Non-degenerate: There exists a $P \in \mathbb{G}_1$ such that $e(P, P) \neq 1$;
- Computability: There exists an efficient algorithm to compute $e(P, Q) \in \mathbb{G}_2$ for all $P, Q \in \mathbb{G}_1$.

We note the modified Weil and Tate pairings associated with supersingular elliptic curves are examples of such admissible pairings. The security of the ID-based multi-signcryption scheme discussed in this paper is based on the following security assumption.

Definition 1. *Given two groups \mathbb{G}_1 and \mathbb{G}_2 of the same prime order q , a bilinear map $e : \mathbb{G}_1 \times \mathbb{G}_1 \rightarrow \mathbb{G}_2$ and a generator P of \mathbb{G}_1 , the Decisional Bilinear Diffie-Hellman problem (DBDHP) in $(\mathbb{G}_1, \mathbb{G}_2, e)$ is to decide whether $h = e(P, P)^{abc}$ given (P, aP, bP, cP) and an element $h \in \mathbb{G}_2$. We define the advantage of a distinguisher against the DBDHP as follows:*

$$AdvD = | P_{a,b,c \in \mathbb{R}Z_q, h \in \mathbb{R}\mathbb{G}_2} [1 \leftarrow D(aP, bP, cP, h)] - P_{a,b,c \in \mathbb{R}Z_q} [1 \leftarrow D(aP, bP, cP, e(P, P)^{abc})] |$$

Definition 2 (Computational Diffie-Hellman (CDH) Assumption)

Let \mathcal{G} be a CDH parameter generator. We say an algorithm \mathcal{A} has advantage $\epsilon(k)$ in solving the CDH problem for \mathcal{G} if for a sufficiently large k ,

$$Adv_{\mathcal{G}, \mathcal{A}}(k) = Pr[\mathcal{A}(q, \mathbb{G}_1, xP, yP) = xyP \mid (q, \mathbb{G}_1) \leftarrow \mathcal{G}^k, P \leftarrow \mathbb{G}_1, x, y \leftarrow \mathbb{Z}_q]$$

We say that the CDH assumption holds in group \mathbb{G}_1 if for any randomized polynomial time algorithm \mathcal{A} , the advantage $Adv_{\mathcal{A}}^{CDH}$ is negligible.

3 Formal Model of ID-Based Directed Signature Scheme without Random Oracle

An ID-based directed signature scheme without random oracle consists of the algorithms $\langle \mathbf{Setup}, \mathbf{Key Extract}, \mathbf{Sign}, \mathbf{Direct Verification}, \mathbf{Public Verification} \rangle$. In the following, we give the detail definitions of their algorithms:

- System parameters initialization (Setup): a private key generator (PKG) generates the system parameters **params** and the master key s to compute the public key $P_{pub} = g^s$, then secretly keep s and make system parameters **params** public.
- Key extraction (Extract): Give an identity ID , PKG computes the private key p_{ID} with his master key s and sends it to the corresponding user through a secret channel.
- Signature generation (Sign): On input a signer's identity ID_s , the designated recipient's identity ID_r , a message M and the private key of signer p_{ID_s} , a signature δ is returned to the designated recipient.
- Directed verification (DVerify): Given the signer's identity ID_s , the designated recipient's identity ID_r and the corresponding signature δ on the message M , the private key p_{ID_r} of the recipient is taken input this algorithm. If the signature δ is valid, then output 1; otherwise, output 0.
- Public verification (PVerify): On input a signer's identity ID_s , the recipient's identity ID_r , a purported signature δ and a third party T , the signer or the designated recipient computes an assistant message **AID**, if δ is valid, output 1, otherwise, output 0.

The security of identity-based directed signature scheme without random oracle consists of two properties: **unforgeability** and **invisibility**.

Unforgeability. For a signature scheme, the well-known strong security notion is existential forgery against chosen message attacks which are proposed by Goldwasser *et al* [1]. Thus, existential unforgeability of ID-based directed signature scheme without random oracle is defined by the following game between a challenger \mathcal{S} and a probabilistic polynomial time attacker \mathcal{A} :

1. \mathcal{S} runs the setup algorithm of ID-based directed signature without random oracle to obtain the public parameters and the master secret. It then gives the public keys and parameters to \mathcal{A} and keeps the master secret itself.
2. Extract query. \mathcal{A} adaptively requests the private key of any identity ID , and \mathcal{B} runs the Extract algorithm on ID to return the private key d_{ID} to \mathcal{A} .
3. Sign query. \mathcal{A} can adaptively query q_s times with input message m_i , the signer's identity ID_A and the verifier's identity's ID_B , and obtains a signature δ_i .
4. Forgery. Finally, \mathcal{A} outputs a signature δ^* for a signer identity ID_A^* , a verifier identity ID_B^* and a message M^* . \mathcal{A} succeeds if the following conditions are satisfied: M^* have been queried for sign oracle with the signer's identity ID_A^* and the verifier's identity's ID_B^* ; the private key of ID_A^* has not been queried on Extract Oracle.

Definition 3. (Unforgeability.) An ID-based directed signature scheme without random oracle is (ϵ, t, q_s, q_e) -unforgeable against chosen message attack and chosen identity attack if there is not polynomial time adversary \mathcal{A} winning the above game with probability greater than ϵ .

Invisibility. The property requires that it should be infeasible for any third party to decide whether a signature on a message m is valid for a signer ID_A and the verifier ID_B . To precisely define this property, we consider the following game between a distinguisher \mathcal{D} and a challenger \mathcal{C} .

Setup. The challenger \mathcal{C} runs **Setup** algorithm with a security parameter k to produce system parameters **params**, then it sends them to the adversary \mathcal{A} and keeps master key secret.

Phase 1. \mathcal{D} performs a series of queries in an adaptive fashion. The following queries are allowed:

Key extraction queries. \mathcal{D} chooses an identity ID . \mathcal{C} computes private key $d_{ID} = \text{Extract}(ID)$ to response to \mathcal{D} .

Signing queries. \mathcal{D} can adaptively query signing oracle q_s times with input message m_i , the signer's identity ID_A and the verifier's identity ID_B , and obtains a signature δ_i .

DVerify queries. \mathcal{D} submits (ID_A, ID_B, m, δ) to \mathcal{C} . The challenger \mathcal{C} first extracts the private key d_B of ID_B , and verify the validity of the signature by this private key. If it is valid, the challenger \mathcal{C} returns 1 to \mathcal{D} , otherwise, it returns 0.

PVerify queries. \mathcal{D} submits (ID_A, ID_B, m, δ) to \mathcal{C} . The challenger \mathcal{C} returns 0 to \mathcal{D} if δ is invalid. Otherwise, the challenger \mathcal{C} produces an assistant message **AID** in the name of the signer or the verifier. Finally, \mathcal{C} sends this **AID** to \mathcal{D} .

Challenge. At the end of phase 1, \mathcal{D} outputs a signer identity ID_A^* , a verifier identity ID_B^* and a message M^* , then it submits them to \mathcal{C} . The constraint is that ID_A^* is not submitted for **Extract oracle**. \mathcal{C} picks a random bit $b \in \{0, 1\}$. If $b = 1$, δ^* is generated as usual using the signing oracle; otherwise, δ^* is chosen uniformly at random from the signature space.

Phase 2. \mathcal{D} can again ask a polynomially bounded number of queries adaptively as in the first phase. But he can make a key extraction query on neither ID_B^* nor ID_A^* . And he also cannot make a DVerify query and a PVerify query on $(\delta^*, ID_B^*, ID_A^*, M^*)$.

Output. \mathcal{D} outputs a bit b' and wins the game if $b' = b$. The advantage of \mathcal{D} is defined as $\text{Adv}_{\mathcal{D}} = |2P[b' = b] - 1|$, where $P[b' = b]$ denotes the probability that $b' = b$.

Two difference between security for ID-based directed signature scheme and conventional directed signature scheme is that firstly an attacker can choose a public identity ID of his choice to attack as opposed to a random public key. Secondly, it is also assumed that the attacker already has some private keys of

some other users in his possession. The definition allows the attacker to obtain a private key associated with any identity of his choice besides the one being attacked.

4 The Scheme

In this section, we give our proposed ID-based directed scheme without random oracle. Our scheme is inspired by Waters’ ID-based encryption scheme. The scheme is described as follows:

Setup. Given k , the PKG chooses two cyclic groups \mathbb{G}_1 and \mathbb{G}_2 of prime order $q > 2^k$, a bilinear pairing map $e : \mathbb{G}_1 \times \mathbb{G}_1 \rightarrow \mathbb{G}_2$ and a generator $g \in_R \mathbb{G}_1$ (Note: for conveniently discuss, we assume that $\mathbb{G}_1 = \mathbb{G}_2$). It then chooses a master key $\alpha \in_R Z_q$, and computes a system-wide public key $P_{pub} = g^\alpha = g_1$. Let H_1 be a cryptographic hash functions where $H_1 : \mathbb{G}_1 \rightarrow \{0, 1\}^*$. Choose $g_2, u, m' \in \mathbb{G}_1$, and two vectors $\mathbf{u} = (u_i)$ and $\mathbf{m} = (m_i)$ of length n_u and n_m respectively. The public parameters are

$$(\mathbb{G}_1, \mathbb{G}_2, P, e, H_1, P_{pub} = g_1, k, g_2, u', m', \mathbf{u}, \mathbf{m})$$

Extract. Let u be a bit string of length n_u , $u[i]$ be the i -th bit of u . Define $\mathcal{U} \subset \{1, 2, \dots, n_u\}$ to be a set of indices i such that $u[i] = 1$. For a user with identity $ID \in \{0, 1\}^*$, its private key is computed as follows:

- Compute $Q_{ID} = H_1(ID) \in \mathbb{G}_1$.
- Randomly choose $k_{ID} \in Z_q$ to compute $p_{ID} = (p_{1_{ID}}, p_{2_{ID}})$, where $p_{1_{ID}} = g_2^\alpha (u' \prod_{i \in \mathcal{U}_{ID}} u_i)^{k_{ID}}, p_{2_{ID}} = g^{k_{ID}}$

Thus, for the sender Alice and the recipient Bob, their private keys are

$$p_{ID_A} = (p_{1_{ID_A}}, p_{2_{ID_A}}) = (g_2^\alpha (u' \prod_{i \in \mathcal{U}_{ID_A}} u_i)^{k_{ID_A}}, g^{k_{ID_A}})$$

and

$$p_{ID_B} = (p_{1_{ID_B}}, p_{2_{ID_B}}) = (g_2^\alpha (u' \prod_{i \in \mathcal{U}_{ID_B}} u_i)^{k_{ID_B}}, g^{k_{ID_B}})$$

Sign. Given a message m , a recipient’s identity ID_B , the sender Alice computes as follows:

1. randomly choose $r_m, k_m \in Z_q$ to compute $\delta_5 = p_{2_{ID_A}} g^{k_m}$ and $\delta_1 = e(g_1, g_2)^{r_m} \oplus \delta_5$
2. compute $\delta_2 = g^{r_m}$
3. compute $\delta_3 = (u' \prod_{i \in \mathcal{U}_{ID_B}} u_i)^{r_m}$.
4. compute $M' = H_1(M)$, where M' is an n_m -bit string and $\mathcal{M}' \subset \{1, 2, \dots, n_m\}$ denotes the set of i for which $M'[i] = 1$.

5. compute $h = H(M||\delta_5||\delta_2)$ and

$$\delta_4 = p_{1_{ID_A}} \left(m' \prod_{j \in \mathcal{M}} m_j \right)^{r_m \cdot h} \left(u' \prod_{i \in \mathcal{U}_{ID_A}} u_i \right)^{k_m}$$

The resultant signature is $\delta = (\delta_1, \delta_2, \delta_3, \delta_4)$.

DVerify. Given a signature $\delta = (\delta_1, \delta_2, \delta_3, \delta_4)$, and a sender's identity ID_A , the recipient with identity ID_B verifies as follows:

1. Firstly, the recipient recovers the value $\delta_5 = \delta_1 \oplus e(p_{2_{ID_B}}, \delta_3)^{-1} \cdot e(p_{1_{ID_B}}, \delta_2)$.
2. compute $M' = H_1(M)$ to obtain the corresponding set \mathcal{M} which is the set of all i for which $M'[i] = 1$.
3. accept the message if and only if the following equation holds

$$e(\delta_4, g) = e(g_1, g_2) e(u' \prod_{i \in \mathcal{U}_{ID_A}} u_i, \delta_5) e(m' \prod_{j \in \mathcal{M}} m_j, \delta_2)^h \quad (1)$$

where $h = H(M||\delta_2||\delta_5)$. If so, output valid; Otherwise, output invalid.

PVerify. Given a signature $\delta = (\delta_1, \delta_2, \delta_3, \delta_4)$, the recipient Bob or the signer Alice computes the aid message $Aid = \delta_1 \oplus e(p_{2_{ID_B}}, \delta_3) \cdot e(p_{1_{ID_B}}, \delta_2)^{-1} = p_{2_{ID_A}} g^{k_m}$ to enable a third party to verify the validity of the signature. Then the signer (or the recipient) sends $\delta' = (\delta_1, \delta_2, \delta_3, \delta_4, \delta_5)$ to the third party in order to verify the validity of this signature by the following equation.

$$e(\delta_4, g) = e(g_1, g_2) e(u' \prod_{i \in \mathcal{U}_{ID_A}} u_i, \delta_5) e(m' \prod_{j \in \mathcal{M}} m_j, \delta_2)^h \quad (2)$$

where $h = H(M||\delta_2||\delta_5)$

5 Security Analysis

We will prove that our proposed scheme is existentially unforgeable and invisible under adaptively chosen-message attack and identity attack in the standard model.

5.1 Correctness

Clearly, the correctness can be easily verified by the following equations.

$$\begin{aligned} & \delta_1 \oplus e(p_{2_{ID_B}}, \delta_3)^{-1} \cdot e(p_{1_{ID_B}}, \delta_2) \\ &= (e(g_1, g_2)^{r_m} \oplus \delta_5) \oplus e(p_{2_{ID_B}}, \delta_3)^{-1} \cdot e(p_{1_{ID_B}}, \delta_2) \\ &= (e(g_1, g_2)^{r_m} \oplus \delta_5) \oplus e(g^{k_{ID_B}}, \delta_3)^{-1} \cdot e(g_2^\alpha (u' \prod_{i \in \mathcal{U}_{ID_B}} u_i)^{k_{ID_B}}, \delta_2) \\ &= \delta_5 \end{aligned}$$

$$\begin{aligned}
e(\delta_4, g) &= e(p_{1_{IDA}} (m' \prod_{j \in \mathcal{M}} m_j)^{r_{m \cdot h}} (u' \prod_{i \in \mathcal{U}_{IDA}} u_i)^{k_m}, g) \\
&= e(g_2^\alpha (u' \prod_{i \in \mathcal{U}_{IDA}} u_i)^{k_{IDA}} (u' \prod_{i \in \mathcal{U}_{IDA}} u_i)^{k_m}, g) e((m' \prod_{j \in \mathcal{M}} m_j)^{r_{m \cdot h}}, g) \\
&= e(g_2^\alpha (u' \prod_{i \in \mathcal{U}_{IDA}} u_i)^{k_{IDA} + k_m}, g) e((m' \prod_{j \in \mathcal{M}} m_j)^{r_{m \cdot h}}, g) \\
&= e(g_2^\alpha, g) e(u' \prod_{i \in \mathcal{U}_{IDA}} u_i, g^{k_{IDA} + k_m}) e((m' \prod_{j \in \mathcal{M}} m_j)^{r_{m \cdot h}}, g) \\
&= e(g_1, g_2) e(u' \prod_{i \in \mathcal{U}_{IDA}} u_i, \delta_5) e(m' \prod_{j \in \mathcal{M}} m_j, \delta_2)^h
\end{aligned}$$

5.2 Security Analysis

In the following, we will show that our scheme satisfies the existential unforgeability and invisibility. Their proofs are given in the full paper [18]

Theorem 1. (Unforgeability) *If a PPT forger \mathcal{A} has an advantage ϵ in forging a signature of ID-based directed signature without random oracle when running in a time t , then (ϵ', t') -CDH assumption can be solved with probability ϵ' .*

Theorem 2. (Invisibility) *If a PPT adversary \mathcal{A} can break the invisibility of our ID-based directed signature scheme with non-negligible probability ϵ , then there exists a distinguisher \mathcal{D} which can solve the DBDH problem with non-negligible probability.*

Efficiency Analysis. Recently, X.Sun *et.al* also proposed a ID-based directed signature scheme [18]. While the security of their scheme is based on random oracle model. In the following, we give a performance comparison of our scheme with Sun *et.al*'s scheme in term of the length of signature, the required computational cost and security. Let C_p be pairing operation, C_e be exponentiation in \mathbb{G}_1 and C_h be hash operation. multiplication operation in \mathbb{G}_1 is neglected. We assume that the bit length of element in \mathbb{G}_1 is $|\mathbb{G}_1|$. Like Waters' signature scheme, we can pre-compute $e(g_1, g_2)$. In the following, the detail comparison is shown in table 1.

From Table 1, we know that our scheme has slightly higher computational cost than Sun *et.al*'s scheme in term of generation and verification of signature. The length of signature in our scheme is more $|\mathbb{G}_1|$ bits than one in the Sun *et.al*'s scheme. Whereas our scheme is proven secure in the standard model. Sun *et.al*'s scheme is only proven secure in the random oracle model. To the best of my knowledge, it is the first scheme which is proven secure in the standard model. All previous scheme mentioned above rely on the random oracle model to prove their security. It is generally believed that cryptographic scheme relying on the random oracles may not be secure if the underlying random oracles are realized as hash function in the real world. For some special applications which require

Table 1. Comparison of our scheme with Sun *et al*'s scheme [18]

Scheme	R.O	Size	DV	PV	Signing cost
Sun <i>et al</i> 's scheme	Yes	$3 \mathbb{G}_1 $	$4C_p + 1C_h$	$4C_p + 1C_h$	$4C_e + 1C_p + 1C_h$
Our scheme	No	$4 \mathbb{G}_1 $	$5C_p + 1C_h$	$5C_p + 1C_h$	$1C_p + 5C_e + 2C_h$

Size denotes the length of signature, DV be designated verification cost, PV be public verification cost, Signing cost be producing signature cost and R.O be whether random oracle is used in the security proof.

very high security, it is believed that only those schemes that can be proven in the standard model must be employed. Thus, our scheme is an efficient and usable scheme.

6 Conclusion

As a special signature, ID-based directed signatures are widely applicable, it is very suitable for special cases in which the designated receiver needs to exclusively verify a signature, and can share the ability to prove validity of the signature others, with the signer. In this paper, we study an ID-based directed signature based on Waters' signature scheme by combining ID-based cryptology and directed signature. And we show that the scheme is secure in the standard model. It satisfies two primitive properties of directed signature: unforgeability and invisibility.

Acknowledgement

This work is supported by Natural Science Foundation of China (NO:60703044), the New Star Plan Project of Beijing Science and Technology (NO:2007B001), the PHR, Program for New Century Excellent Talents in University (NCET-06-188), The Beijing Natural Science Foundation Programm and Scientific Research Key Program of Beijing Municipal Commission of Education (NO:KZ2008 10009005) and 973 Program (No:2007CB310700).

References

1. Chaum, D., van Antwerpen, H.: Undeniable Signatures. In: Brassard, G. (ed.) CRYPTO 1989. LNCS, vol. 435, pp. 212–216. Springer, Heidelberg (1990)
2. Lu, R., Lin, X., Cao, Z., Shao, J., Liang, X.: New (t, n) threshold directed signature scheme with provable security. Information Sciences 178, 756–765 (2008)
3. Lu, R., Zhen, F., Zhou, Y.: Threshold undeniable signature scheme based on conic. Applied mathematics and computation 162, 165–177 (2005)
4. Chaum, D.: Designated Confirmer Signatures. In: De Santis, A. (ed.) EUROCRYPT 1994. LNCS, vol. 950, pp. 86–91. Springer, Heidelberg (1995)
5. Jakobsson, M., Sako, K., Impagliazzo, R.: Designated Verifier Proofs and Their Applications. In: Maurer, U.M. (ed.) EUROCRYPT 1996. LNCS, vol. 1070, pp. 143–154. Springer, Heidelberg (1996)

6. Lim, C.H., Lee, P.J.: Modified Maurer-Yacobi's scheme and Its application. In: Advances in Cryptology-ACISP 1992. LNCS, vol. 718, pp. 308–323. Springer, Heidelberg (1992)
7. Laguillaumie, F., Paillier, P., Vergnaud, D.: Universally convertible directed signatures. In: Roy, B. (ed.) ASIACRYPT 2005. LNCS, vol. 3788, pp. 682–701. Springer, Heidelberg (2005)
8. Bellare, M., Rogaway, P.: Random oracles are practical: a paradigm for designing efficient protocols. In: Proceedings of the First Annual Conference on Computer and Communications Security, pp. 62–73. ACM Press, New York (1993)
9. Sunder, L., Manoj, K.: A Directed Threshold-Signature Scheme, <http://arxiv.org/ftp/cs/papers/0411/0411005.pdf>
10. Shamir, A.: How to share a secret. Communications of the ACM 22, 612–613 (1979)
11. Schnorr, C.P.: Efficient signature generation by smart cards. Journal of Cryptology 4, 161–174 (1994)
12. Shamir, A.: Identity-Based Cryptosystems and Signature Schemes. In: Blakely, G.R., Chaum, D. (eds.) CRYPTO 1984. LNCS, vol. 196, pp. 47–53. Springer, Heidelberg (1985)
13. Bellare, M., Neven, G.: Identity-based Multi-signatures from RSA. In: Abe, M. (ed.) CT-RSA 2007. LNCS, vol. 4377, pp. 145–162. Springer, Heidelberg (2006)
14. Libert, B., Quisquater, J.J.: Identity based undeniable signatures. In: Okamoto, T. (ed.) CT-RSA 2004. LNCS, vol. 2964, pp. 112–125. Springer, Heidelberg (2004)
15. Waters, B.: Efficient Identity-based encryption without random oracles. In: Cramer, R. (ed.) EUROCRYPT 2005. LNCS, vol. 3494, pp. 114–127. Springer, Heidelberg (2005)
16. Boneh, D., Lynn, B., Shacham, H.: Short signature from the Weil pairing. Journal of Cryptology 17, 297–319 (2004)
17. Boneh, D., Boyen, X.: Short signatures without random oracles. In: Cachin, C., Camenisch, J.L. (eds.) EUROCRYPT 2004. LNCS, vol. 3027, pp. 56–73. Springer, Heidelberg (2004)
18. Goldwasser, S., Micali, S., Rivest, R.: A digital signature scheme secure against adaptively chosen message attacks. SIAM Journal on Computing 17, 281–308 (1998)

Mask Particle Filter for Similar Objects Tracking

Huaping Liu^{1,2}, Fuchun Sun^{1,2}, and Meng Gao³

¹ Department of Computer Science and Technology, Tsinghua University,
Beijing 100084, China

² State Key Laboratory of Intelligent Technology and Systems,
Beijing 100084, China

³ Shijiazhuang Railway Institute, Shijiazhuang 050043, China

Abstract. Tracking appearance similar objects is very challenging. Conventional approaches often encounter “hijack” problem. That is to say, the tracking results for the smaller objects will be attracted to the larger one in the close vicinity. In this paper, we propose a decentralized particle filter approach for similar objects tracking. When the objects are close, the tracking results for the larger one will be masked and its influence will be eliminated. In principle, the tracker for the smaller object needs to be run two times, which increase the time costs. To tackle this, we construct the integral image for the mask region and dramatically decrease the calculation time of the evaluation of likelihood functions in the masked image. Experimental results show that the proposed approach effectively avoids “hijack” problems.

Keywords: Visual tracking, Particle filter.

1 Introduction

Multiple object tracking in video sequences is a wide explored topic with a great number of applications, such as security, military tasks or traffic control [1,2]. However, tracking of similar objects in appearance often fails when they are in close proximity or present occlusions, since in this case they cannot be treated independently. In such circumstances, multiple independent single object trackers suffer from the well-known “hijack” (also called “error merge”) problem, which results that some trackers lose their associated objects and falsely coalesce with other objects.

One of the most famous over the last decades was the particle filter, a probabilistic tracking approach which can deal with nonlinear and non-Gaussian problem. To tackle the “hijack” problem when using independent particle filters for multiple objects tracking, Ref.[3] proposed a joint particle filter which uses a Markov random field to model motion interaction. Ref.[4] pointed that although this framework is powerful, its current implementation cannot handle severe occlusions (we can also see this phenomenon in the experimental parts of this paper). In addition, joint particle filter like [3] requires a tremendous computational cost due to the complexity introduced by the high dimensionality of the joint state representation. The complexity of most implementations based

on a joint state-space representation grows exponentially in terms of the number of objects tracked. Recently, [4] proposed a decentralized particle filter approach for multiple object tracking. This approach used a magnetic-inertia potential model to model the interaction between objects and implicitly handle “error merge” problem. However, the modelling for the interaction between objects is complicated.

In this paper, we also use the decentralized particle filter scheme. A novelty of this approach is that the modelling for interaction between objects is not needed. The idea is borrowed from humans visual cognition. When people watch some scene and focus on some object, he usually neglects the surroundings. By this means, he can persistently capture small objects, while there are large similar objects around it. The kernel of the proposed approach is to mask the region of tracked results and then re-track region-of-interest. In this way, the tracker can suppress the disturbance from surroundings. We call this approach as Mask Particle Filter. Though the idea is straightforward, the algorithm requires careful designs. In the second section, we will give a very brief introduction about conventional particle filter and present the proposed approach. Section 3 gives the comparison results.

2 Mask Particle Filter

Particle filter is one of the most used tracker and therefore we first briefly review the conventional particle filter.

The task of tracking is to use the available measurement information to estimate the hidden state variables. Given the available observations $\mathbf{z}_{1:k-1} = \mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_{k-1}$ up to time instant $k-1$, the prediction stage utilizes the probabilistic system transition model $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ to predict the posterior at time instant k as

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})d\mathbf{x}_{k-1} \quad (1)$$

At time instant k , the observation \mathbf{z}_k is available, the state can be updated using *Bayes's* rule

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) = \frac{p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k-1})}{p(\mathbf{z}_k|\mathbf{z}_{1:k-1})} \quad (2)$$

where $p(\mathbf{z}_k|\mathbf{x}_k)$ is described by the observation equation.

In general, the integrals in (1) and (2) are analytically intractable. To solve this problem, the particle filter approaches are proposed [5]. The kernel of particle filter is to recursively approximate the posterior distribution using a finite set of weighted samples. Each sample \mathbf{x}_k^i represents one hypothetical state of the object, with a corresponding discrete sampling probability ω_k^i , which satisfies $\sum_{i=1}^N \omega_k^i = 1$. The posterior $p(\mathbf{x}_k|\mathbf{z}_{1:k})$ then can be approximated as

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) \approx \sum_{i=1}^N \omega_k^i \delta(\mathbf{x}_k - \mathbf{x}_k^i) \quad (3)$$

where $\delta(\cdot)$ is Dirac function.

The candidate samples $\{\mathbf{x}_k^i\}_{i=1,2,\dots,N}$ are drawn from a proposal distribution $q(\mathbf{x}_k|\mathbf{x}_{1:k-1}, \mathbf{z}_{1:k})$ and the weight of the samples are

$$\omega_k^i = \omega_{k-1}^i \frac{p(\mathbf{z}_k|\mathbf{x}_k^i)p(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i)}{q(\mathbf{x}_k|\mathbf{x}_{1:k-1}, \mathbf{z}_{1:k})} \quad (4)$$

The samples are re-sampled to generated an unweighed particle set according to their importance weights to avoid degeneracy. In the case of the bootstrap filter [3], $q(\mathbf{x}_k|\mathbf{x}_{1:k-1}, \mathbf{z}_{1:k}) = p(\mathbf{x}_k|\mathbf{x}_{k-1})$ and the weights become the observation likelihood $p(\mathbf{z}_k|\mathbf{x}_k)$.

In visual tracking, the color histogram is an extensively used feature [6]. Color distributions are used as object models as they achieve robustness against non-rigidity, rotation and partial occlusion. In our experiments, the histograms are typically calculated in the RGB space using $8 \times 8 \times 8$ bins. The resulting complete histogram is thus composed of $N_h = 512$ bins.

The color-similarity measure is based on the similarity between the color histogram of a reference region and that of the image region in frame k represented by a sample \mathbf{x}_k^i . To estimate the proper weight for this sample during the measurement update step, we need the observation model $p(\mathbf{z}_k|\mathbf{x}_k = \mathbf{x}_k^i)$. This model can be obtained by the following equation

$$p(\mathbf{z}_k|\mathbf{x}_k = \mathbf{x}_k^i) \propto \exp\{-\lambda D^2(\mathbf{q}^*, \mathbf{q}_k(\mathbf{x}_k^i))\} \quad (5)$$

where λ is an experimentally determined constant and \mathbf{q}^* and $\mathbf{q}_k(\mathbf{x}_k^i)$ are the color histograms of the reference region and the region defined by \mathbf{x}_k^i , respectively. The distance measure $D(\cdot, \cdot)$ is derived from the Bhattacharyya similarity coefficient and is defined as

$$D^2(\mathbf{q}^*, \mathbf{q}_k(\mathbf{x}_k^i)) = \{1 - \sum_{n=1}^{N_h} \sqrt{q^*(n)q_k(n; \mathbf{x}_k^i)}\}^{1/2} \quad (6)$$

More details can be found in [7]. Although color histogram is a robust feature, it also present some disadvantages.

In visual tracking, there usually happens overlapping between objects. When the tracking boxes of two objects are overlapped each other, there will be two possibilities:

(1) One of the objects is indeed occluded by the other object;

(2) Both of the objects are not overlapped. In this case, the tracking results of the smaller object is “hijacked” by the larger one. If “hijack” happens, the tracker usually cannot recover good performance even if these object move apart.

To use the mask particle filter, we should tackle some key problems. The first task is to determine which object should be masked. Therefore, when there are



Fig. 1. Mask image

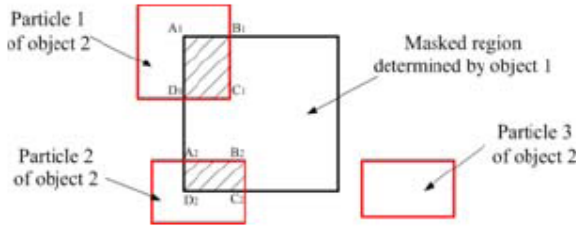


Fig. 2. Integral image for mask region

overlaps between tracking boxes, we need launch an **Occlusion Analysis** module. This module is used to determine which tracking box should be masked. In general saying, any occlusion analysis approach can be used in this stage. In our work, we use the similarity with the corresponding reference histogram to evaluate it. The tracking box with higher similarity, which means that it is accurately tracked and may hijack other objects, should be masked. For example, in the left of Fig.1, we obtained two tracking boxes, which are overlapped; therefore we calculate their similarity with their reference histogram respectively. Since the white box (we assume it to be object 1) has higher similarity, it is masked (see the right of Fig.1). For convenience, we call this masked image as **Masked Image**, which is an important intermediate image.

Since object 1 is masked, its influence has been eliminated. In the following, we should again run the particle filter for object 2 to get its tracking results (see the right of Fig.1). This is very straightforward but very time-consuming, since the particle filter for object 2 will be run two times and the time costs will be doubled. Recall that the particle filter can be mainly divided into three stages: prediction, likelihood computation and re-sampling, where likelihood computation is the most time-consuming. Fortunately, we notice that the particle filter in current frame for object 2 has been run once and therefore the obtained prediction particles (which has not been re-sampled) can be used in the second run of the particle filter. Therefore the prediction stage can be omitted. After then, all of the prediction particles should be evaluated in the Masked Image, but not the original image. If we directly calculate their histogram on the Masked Image, the time-cost will be high. To remedy it, we propose an improved approach.

If some particle (such as Particle 3 in Fig.2) is not overlapped with the masked region, then its histogram will not be altered. In fact, many particles of object 2 will be overlapped with the mask region, since object 1 and object 2 are rather close. Because we have obtained the histograms of the particles in the original image, we just use them to subtract the histogram of the overlapped region (see rectangles $A_1B_1C_1D_1$ or $A_2B_2C_2D_2$ in Fig.2), then the histograms of these particles in the **Masked Image** can be obtained. For speeding the computation, we construct an integral image of the histogram of the masked region (but **NOT** the whole image). The integral image is an intermediate image representations used for fast calculation of region sums. Each pixel of the integral image is the sum of all the pixels inside the rectangle bounded by the upper left corner of the image and the pixel of interest. In [7], the integral image was extended to higher dimensions for fast calculation of region histograms. Here we adopt similar technology to re-evaluate the new likelihoods of the particles in the **Masked Image**. After the integral image is constructed, we just need some simple calculation to obtain the new histograms of the particles. This will dramatically reduce the likelihood evaluation time. Finally, we use the new likelihood functions to produce the weights of the particles and use these weights to resample the particles. Notice that we just construct the integral image for the masked region, but not the whole image. In addition, the re-sampling is only needed in the second run of the particle filter for object 2.

The above-mentioned approach can effectively deal with cases where two object are close, even in occlusion. In addition, if object 2 is indeed occluded by object 1, then the similarity of the estimated results for object 2 will be small, since the masked region of object 1 also masked some part of object 2. We can use some mechanisms to judge whether object 2 is occluded or not.

If object 2 is occluded, we will give up tracking it. Instead, we persistently track object 1 and uniformly sampling particles for object 2 around the estimated position of object 1. This mechanism will help to recapture object 2 when it reappears.

The main procedure of the proposed algorithm can be briefly summarized as follows ($i = 1, 2$ is the label of objects. In this algorithm we assume object 1 may hijack object 2, which is without loss of generality):

(1) For each objects, independent particle filters $\mathbf{PF}_i(\cdot)$ is used to obtain the prediction particles with weights $[\bar{\mathbf{x}}_{i,k}^{(1:M)}, \bar{\omega}_{i,k}^{(1:M)}] = \mathbf{PF}_i(\mathbf{Image}_k, x_{i,k-1}^{(1:M)})$, where \mathbf{Image}_k is the original frame at time k .

(2) Obtain the estimation $\hat{\mathbf{x}}_{i,k} = \sum_{j=1}^M \bar{\mathbf{x}}_{i,k}^j \bar{\omega}_{i,k}^j$.

(3) IF $\hat{\mathbf{x}}_{1,k}$ and $\hat{\mathbf{x}}_{2,k}$ are in close vicinity, then calculate the similarity of each estimation $S_i = \mathbf{Similarity}(\hat{\mathbf{x}}_{i,k})$.

(4) IF ($S_1 > S_2$) THEN obtain Mask_Image_k from Image_k by masking the region determined by $\hat{\mathbf{x}}_{1,k}$.

(5) Calculate the integral histogram for the masking region determined by $\hat{\mathbf{x}}_{1,k}$.

(6) Recalculate the histogram for each particle $\bar{\mathbf{x}}_{2,k}^j$, and the corresponding weight value $s \omega_{2,k}^j$.

(7) Modify the estimation of object 2 as $\hat{\mathbf{x}}_{2,k} = \sum_{j=1}^M \bar{\mathbf{x}}_{2,k}^j \omega_{2,k}^j$.

(8) Re-sample $(\bar{\mathbf{x}}_{1,k}^{(1:M)}, \bar{\omega}_{1,k}^{1:M})$ to get $[\mathbf{x}_{1,k}^{(1:M)}]$; and re-sample $(\bar{\mathbf{x}}_{2,k}^{(1:M)}, \bar{\omega}_{2,k}^{1:M})$ to get $[\mathbf{x}_{2,k}^{(1:M)}]$.

3 Experimental Results

In this section, we demonstrate our algorithm with a practical video for two similar cars tracking, where object 1 (bounded by red box) moves from right to left, and object 2 (bounded by yellow box) moves from left to right. Therefore they present strong overlaps during Frames 320-365. The video sequence consists of 395 frames at 640×480 pixels resolution.

We have compared the performances of the proposed mask particle filter with Khan's particle filter [3]. In Khan's particle filter, we construct the Markov random field constraints as $\phi_{i,j} = \exp(-\lambda \mathbf{N}(\mathbf{x}^i, \mathbf{x}^j))$ (see [8] for its details), where $\mathbf{N}(\mathbf{x}^i, \mathbf{x}^j)$ represents the number of the overlapped region between two areas determined by two particles \mathbf{x}^i and \mathbf{x}^j . We find that Khan's approach is sensitive to the parameter λ . When λ is too small, the effect of Markov random field constraints is weak and this approach may reduce to conventional independent particle filters. When λ is too large, the repulsive force between particles is too high and therefore it is difficult for the resulting approach to deal with cases where two objects are in close vicinity. How to select this parameter is still an open problem. In this paper, we experimentally verified a lot of selections of λ and cannot find a suitable value. It seems that $\lambda \in [0.0001, 0.0005]$ is a little suitable. However, none of any fixed value can adapt to all cases throughout this video.

For all of the experiments, the state of the each particle filter is defined as $\mathbf{x}_k = [x_k, y_k, s_k]$, where x_k, y_k indicate the location of the object, s_k the scale. The dynamics of the objects are assumed to be a random walking model, which can be represented as $\mathbf{x}_k = \mathbf{x}_{k-1} + \mathbf{v}_k$, where \mathbf{v}_k is a multivariate zero-mean Gaussian random variable. Its variances are set by $[\sigma_x, \sigma_y, \sigma_s] = [10, 10, 0.05]$. For each particle filter, we assign 100 samples. In Khan's approach, since the joint particle filter is utilized, we assigned 200 particles for the joint particle filter. In addition, since there are only two objects, we used conventional importance sampling instead of MCMC sampling in [8]. We think this does not alter the intrinsics of Khan's algorithm.

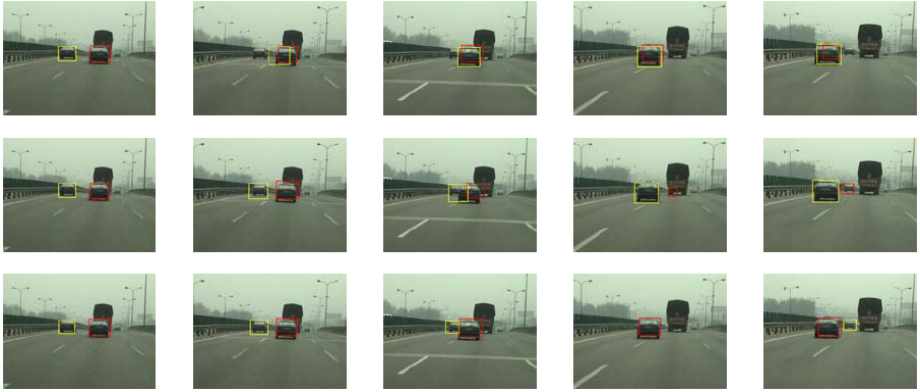


Fig. 3. Tracking results comparison. The first row: Khan’s approach ($\lambda=0.0001$); The second row: Khan’s approach ($\lambda = 0.0005$); The third row: The proposed mask particle filter. From left to right: Frame 275, Frame 300, Frame 325, Frame 350, Frame 375.

For fair comparison, all of the particle filters for the sequence are started with same initial detection results, which are manually labelled. During likelihood evaluation, we use fusion of RGB color histogram and edge orientation histogram.

Fig.3 presents some representative examples. We can see during initial stage (From start to Frame 275 or so), all of the particle filters work well, since both objects are far. At Frame 300, we see that Khan’s approach with $\lambda = 0.0001$ fails to track object 1(in yellow). The reason is that the parameter λ is too small and the MRF constraints does not work. We made a lot of experiments and found that for any $\lambda < 0.0005$, the tracking results are similar. A worse result is that both trackers mistakenly merge together after Frame 300.

When $\lambda = 0.0005$, Khan’s approach works well at Frame 300, which means that MRF constrains play roles when the objects are in close vicinity. However, when both objects moves more near (see Frame 325), This approach cannot give satisfactory results, since the MRF constraints is too strong and the approach can not deal with these cases. After Frame 325, we see from the second row that the tracker of object 2 occupies the position of object 1, while the tracker of object 1 is forced to an error position. In Frame 375, there even appears a totally error association result.

In the third row, we can see the results of our tracking algorithm. This algorithm work well from Frames 300-325, in which the objects are more and more close.

In addition, in frame 350, object 1 totally occludes object 2. Object 1 is accurately tracked and then its region is masked. In the masked image, tracker 2 can be determined. However, its similarity with its reference histogram is too low and therefore it is determined as being occlusion, during which the tracker for object 2 finishes tracking, but uniformly sample particles for object 2 around object 1 to re-capture object. In practical video, object re-appears from Frame 368, and our algorithm succeeds re-capturing it from Frame 375.

4 Conclusions

An important merit of the proposed approach is that explicit modelling for interaction between objects is not required, which can not be avoided by [5] and [3]. The presented examples restrict to two objects with similar appearances, but the approach can be easily extended to multiple objects case. The proposed approach of mask particle filter is a general framework, it does not make any assumptions about the individual tracker. Therefore any other tracking approach can be incorporated into it. In addition, some other advanced occlusion reasoning approaches can be adopted to enhance the robustness. Our current work is continuing to develop mask particle filter for multiple object tracking in a more formal sense, and to apply these insights to challenging tracking problems.

Acknowledgements. This work was jointly supported by the National Science Fund for Distinguished Young Scholars (Grant No. 60625304), the National Natural Science Foundation of China (Grants No. 90716021, 60621062, 60572178), the National Key Project for Basic Research of China (Grants No. G2007CB311003, 2009CB724002), and National High-tech Research and Development Program (2007AA04Z232).

References

1. Alefs, B., Schreiber, D., Clabian, M.: Hypothesis based vehicle detection for increased simplicity in multi sensor ACC. In: Proc. of IEEE Intelligent Vehicles Symposium, pp. 261–266 (2005)
2. Dellaert, F., Thorpe, C.: Robust car tracking using Kalman filtering and Bayesian templates. In: SPIE Conference on Intelligent Transportation Systems, pp. 72–83 (1997)
3. Doucet, A., De Freitas, N., Gordon, N.: Sequential Monte Carlo Methods in Practice. Springer, New York (2001)
4. Du, M., Guan, L.: Monocular human motion tracking with the DE-MC particle filter. In: Proc. of Int. Conf. on Acoustics, Speech, and Signal Processing, pp. 205–208 (2006)
5. Fu, C., Huang, C., Chen, Y.: Vision-based preceding vehicle detection and tracking. In: Proc. of Int. Conf. on Pattern Recognition, pp. 1070–1073 (2006)
6. Xue, J., Zheng, N., Zhong, X.: An integrated monte carlo data association framework for multi-object tracking. In: Proc. of Int. Conf. on Pattern Recognition, pp. 703–706 (2006)
7. Khan, Z., Balch, T., Dellaert, F.: MCMC-based particle filtering for tracking a variable number of interacting targets. IEEE Trans. on Pattern Analysis and Machine Intelligence 27, 1805–1819 (2005)
8. Qu, W., Schonfeld, D., Mohamed, M.: Real-time distributed multi-object tracking using multiple interactive trackers and a magnetic-inertia potential model. IEEE Transactions on Multimedia, 511–519 (2007)
9. Perez, P., Hue, C., Vermaak, J., Gangnet, M.: Color-based probabilistic tracking. In: Proc. of European Conf. on Computer Vision, pp. 661–675 (2002)

10. Porikli, F.: Intergral histogram: A fast way to extract histograms in Cartesian spaces. In: Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition, pp. 829–836 (2005)
11. Hamlaoui, S., Davoine, F.: Facial action tracking using an AAM-based condensation approach. In: Proc. of Int. Conf. on Acoustics, Speech, and Signal Processing, pp. 701–704 (2005)
12. Hilario, C., Collado, J.M., Armingol, J.M., De La Escalera, A.: Pyramidal image analysis for vehicle detection. In: Proc. of IEEE Intelligent Vehicles Symposium, pp. 88–93 (2005)
13. Maggio, E., Cavallaro, A.: Hybrid particle filter and mean shift tracker with adaptive transition model. In: Proc. of Int. Conf. on Acoustics, Speech, and Signal Processing, pp. 221–224 (2005)
14. Okuma, K., Taleghani, A., De Freitas, N., Little, J.J., Lowe, D.G.: A boosted particle filter: multitarget detection and tracking. In: Proc. of European Conf. on Computer Vision, pp. 28–39 (2004)
15. Schweiger, R., Neumann, H., Ritter, W.: Multiple-cue data fusion with particle filters for vehicle detection in night view automotive applications. In: Proc. of IEEE Intelligent Vehicles Symposium, pp. 753–758 (2005)
16. Zielke, T., Brauckmann, M., Seelen, W.V.: CARTRACK: Computer vision-based car-following. In: Proc. of IEEE Workshop on Applications of Computer Vision, pp. 156–163 (1992)

An Efficient Wavelet Based Feature Extraction Method for Face Recognition

Iman Makaremi and Majid Ahmadi

Department of Electrical and Computer Engineering,
University of Windsor,
Windsor N9B 3P4, Canada
{makarem, ahmadi}@uwindsor.ca

Abstract. A computationally efficient wavelet based feature extraction method is proposed. This method is used for face recognition along with an HMM classifier. In comparison to similar method, this method needs less computation while the highest possible classification rate is still achievable. In this paper, different wavelet filters have been tried and effect of sub-image's size and overlap percentage in feature extraction on classification rate has been studied.

Keywords: Face recognition, Feature extraction, Wavelet analysis, Hidden markov models.

1 Introduction

Face recognition is still an exciting field to researchers in spite of hundreds of publications in that field. Moreover, the need for having a 100% or at least an almost flawless algorithm is still in demand. Also, it will be more attractive if the devised algorithm is easy and cheap to implement since demand for security systems has been growing exponentially in the last decade, and face recognition has become the most important part of such systems.

Many different techniques have been introduced and applied to face recognition. Geometric features based approaches [1], Eigenface [2], Independent Component Analysis (ICA) [3], and Elastic Matching method [4] are the most well-established methods which have been used so far.

In addition, Hidden Markov Model (HMM) [5] is proven to be an effective technique for face recognition. To represent faces as observation sequences to HMM, different methods of feature extraction have been proposed. Discrete Cosine Transform (DCT) [6], gray tone features [7], and Discrete Wavelet Transform (DWT) are few to mention. The combination of DWT as the feature extractor and HMM as the classifier has resulted in yielding very satisfactory outcome compared to other techniques for face recognition. For example, Otsuka et al. [8] applied wavelet transform to sub-images of faces up to five levels and the average of power and phase of the wavelet coefficients are used as observation sequences. Then, HMM is used as the facial expression recognizer. They used it on a database which was produced by them and contained different expression of three male subjects. The recognition rate was up to 98 percent. Scanning the face with a fixed square window on a curvy path is the

proposed method by Bicego et al. [9]. After calculating the wavelet features of each sub-image, some of the wavelet coefficients are retained and used as the features. They also used HMM as the classifier and applied it on AT&T database. The classification rate was 100 percent. Hung-Son et al. [10] presented a different way of feature extraction which is a horizontal followed by a vertical scanning and then calculating the wavelet coefficients. The noticeable part of this method is using only one HMM for classification which is proven to increase the recognition speed. However, this requires more memory for storing information about the classes in the database. Nicholl et al. [11] used 2DWT for feature extraction along with Structural Hidden Markov Model (SHMM) [12] as the classifier. For feature extraction, faces are decomposed into blocks and the L_2 norm of wavelet coefficients of blocks are used as feature vectors. The remarkable part of this method is combining the multiresolution features with the local interactions of the facial structures expressed through the SHMM.

In this paper, we introduce a simpler technique for feature extraction using wavelet transform. This method requires less computation which makes it more attractive when compared with similar methods in the literature. In the next section, the proposed method is described in detail. Section III deals with continuous HMM and its application as a classifier. In section IV, experimental results and comparison are presented. Finally, conclusions are presented in section V.

2 Wavelet Feature Extraction

Wavelet transform [13] has been used widely in many fields such as JPEG2000 [14]. A growing number of publications deal with hardware implementation of this transform [15-18] which demonstrate its utility in the area of DSP and Pattern Recognition. A multi-resolution analysis of a signal with localization in both time and frequency is the advantage of wavelet transform over Fourier and cosine transforms [13, 19]. Also, having different alternatives for the basis function in wavelet transform makes it more adaptable for different problems than Fourier transform since in the latter only one kind of basis function can be used.

In order to perform feature extraction, the 2D wavelet transform of the faces is calculated using different basis functions only up to one level. Having four sets of coefficient matrices, one as the approximation and the other three as details (horizontal, vertical, and diagonal) for each face, the features are extracted as follows.

A window with a width which is equal to the width of the coefficient matrices and an arbitrary height k is selected. Superimposing it on top of the matrices, mean, variance, and absolute mean of the content of the window is calculated. Since there are four matrices and three features for the window, there are 12 features for that section of the wavelet transform. Then, the window is slid down. Here, an overlap factor can be defined such that each two neighboring windows can overlap by r rows (Fig. 1). The whole process is carried out till the window reaches to the bottom of the image. Finally, there will be a sequence with 12 features which represents each face. The length of the sequence depends on three parameters; 1) basis function used for wavelet transform, 2) the height of the window k , and 3) the overlap r . In this work, we used two different filters of different orders; Daubechies (abbreviated as DB in the

tables) of orders one (which is equivalent to Haar) and ten, and Coiflet (abbreviated as COIF in the tables) of orders one and two. Also two different values for k and three different values for r have been considered.

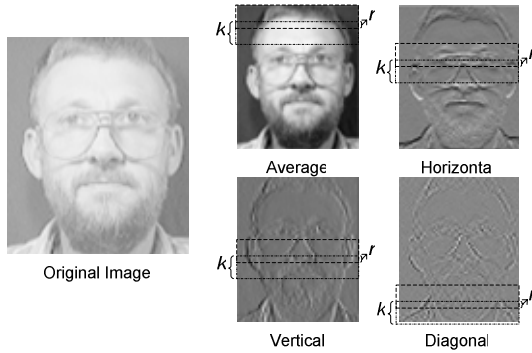


Fig. 1. A wavelet decomposition (Haar filter is used for this representation) and the top to bottom scanning for feature extraction. The wavelet coefficients have been normalized between 0 and 1 for representation (Images are resized for representation).

To study the class separability, the *within-class* and *mixture scatter* matrices have been calculated. Within-class matrix is calculated as follows

$$S_w = \sum_{i=1}^K P_i S_i$$

where S_i is the covariance matrix of i 'th class and P_i is the *a priori* probability of the corresponding class. Also, mixture scatter matrix is the covariance matrix of the whole data calculated as follows:

$$S_m = E[(x - u_0)(x - u_0)^T]$$

where u_0 is the mean of all classes. The class separability criterion which is used in this work is $J = trace(S_m) / trace(S_w)$. J 's have been rounded and presented in Table 1 in different categories based on filters, k , and r . Based on the results presented in Table 1, by increasing the height of the window, the class separability decreases. Also, for a fixed height of the window, the class separability increases with increasing the overlap.

The proposed method which can be implemented easily requires less computation than any other methods cited in the literature. The reason of this claim is that wavelet decomposition is applied on the whole image and for only one level, while in other methods it is applied on each sub-image separately which mostly has overlaps with its neighbors [9, 10, and 11] and in some cases, the wavelet transform is calculated up to five level [8]. This normally translates to more computation for wavelet decomposition. Experimental results show that, in MATLAB environment, it takes 0.04 seconds to extract features for each face with a 64X2 2.41GHz/Win XP.

Table 1. Class separability measures with different filters, k 's, and r 's

$k \backslash r$		1	2	3		1	2	3
4	COIF1	348	570	1130	DB1	375	553	1096
5		295	386	573		283	372	556
4	COIF2	402	600	1179	DB10	438	644	1262
5		304	403	595		327	430	636

3 Hidden Markov Model as Classifier

HMMs as double stochastic process can be used to characterize the statistical properties of signals [5]. In fact, a signal is considered as a sequence of observation which can be observed directly. There are basically two different kinds of observations, discrete and continuous. In this paper, we use the continuous HMM since our extracted sequences from faces are continuous. Furthermore, it is not recommended to discretize the output as long as it is possible [5]. A continuous HMM λ is defined by the elements as follow:

- Q , the number of hidden states in the model
- T , length of sequences
- $\delta = \{S_1, \dots, S_Q\}$, the finite set of possible hidden states.
- $\Pi = \{\pi_i\}$, the initial state probability distribution, where, $\pi_i = P[q_1 = S_i], 1 \leq i \leq Q$ and $\sum_{i=1}^Q \pi_i = 1$.
- $A = \{a_{ij}\}$, the state transition probability matrix, where $a_{ij} = P[q_{t+1} = S_j | q_t = S_i], 1 \leq i, j \leq Q$ and $\sum_{j=1}^Q a_{ij} = 1, 1 \leq i \leq Q$.
- $B = \{b_{j,t}\}$ the emission probability matrix,
- where $b_{j,t} = P[O_t | q_t = S_j], 1 \leq j \leq Q, 1 \leq t \leq T$. There are different approaches to define the emission probability for continuous observations. The most general representation of the PDF is a finite mixture of the form $b_{j,t} = \sum_{m=1}^M c_{jm} N(O_t, u_{jm}, U_{jm}), 1 \leq j \leq Q$ where c_{jm} , the mixture coefficient for the m th mixture in state j is always greater than or equal to zero and summation over m should be equal to 1. N is a Gaussian function, and u_{jm} and U_{jm} are the mean vector and the covariance matrix of the m th mixture component in state j respectively.

To have a functional HMM for real-world applications, three basic problems should be solved. These problems are

- **Evaluation:** Calculating $P(O | \lambda)$
- **Decoding:** Choosing the state sequence that explains the observations.
- **Parameter Estimation:** Adjusting the model parameters.

HMMs can be used as classifiers in two different ways; path discriminant and model discriminant [20]. In path discriminant approach, only one HMM is used for all classes and different state sequences of the model distinguish classes. While in the model discriminant approach, a separate model is used for each class and the class label is obtained based on the probability of output:

$$c = \arg \max_{1 \leq i < L} [P(O | \lambda_i)]$$

where, L is the total number of classes. In this paper, we used the second approach where a distinct model is built for each individual class. We use Baum-Welch method [5] for training and considering that there are more than one sample in training set for each class, the modified version of this method is utilized [21].

4 Results, Comparison, and Discussion

AT&T (already known as ORL) is the database we used in this paper. This database includes 400 different pictures of 40 individuals, 10 for each. Five out of ten photos were randomly put in the training set and the rest were put in the testing set. Different numbers of states and mixture components have been tried to study their effects on classification rates. Therefore, models with 2, 5, and 8 states and 2 and 4 mixture components were tried on all of the different feature sets extracted in the last section. To realize the effect of the number of hidden states on the classification rates, all of the models have been sorted based on their classification rates and the best n models have been chosen and shown in Table 2 based on different n and the number of their states, e.g. between the ten best models, two of them have 2 hidden states, seven have 5 hidden states and just one has 8 hidden states. Based on the information in Table 2, 5-hidden state models generally performed better than the other two.

Table 2. The influence of number of hidden states (H.S.) on classification rate on test set

No. H. S.	Top 10	Top 50	Top 100	Top 200
2	2	12	24	40
5	7	23	39	90
8	1	15	37	70

The same study has been conducted for the effect of the number of mixture components. Table 3 shows the results that models with 2 mixture components are generally performing better.

Table 3. The influence of mixture components (M. C.) on classification rate on test set

No. M. C.	Top 10	Top 50	Top 100	Top 200
2	9	38	62	110
4	1	12	38	90

Tables 4 and 5 show the average and the maximum classification rates of a 5-state 2-mixture component model respectively. The maximum average classification rate is 0.988 which means less than 4 faces were misclassified. This was achieved when the wavelet filter was Daubechies of order one which is the simplest wavelet basis. Table 5 shows that the highest classification rate is 1 where no misclassification happened and it is achieved by a Coiflet filter of order one. With other filters, results were not extremely different. The maximum classification rate that was achieved for the others is 0.995 where only one face is misclassified.

Table 4. Average classification rate on test set with different filters, k 's, and r 's

$k \backslash r$	1			2			3		
4	COIF1	0.957	0.973	0.983	DB1	0.957	0.975	0.988	
5		0.968	0.962	0.964		0.981	0.982	0.985	
4	COIF2	0.959	0.944	0.958	DB10	0.964	0.978	0.951	
5		0.953	0.973	0.947		0.976	0.978	0.971	

Table 5. Maximum classification rate on test set with different filters, k 's, and r 's

$k \backslash r$	1			2			3		
4	COIF1	0.975	0.990	0.995	DB1	0.990	0.980	0.995	
5		0.980	0.980	0.975		0.990	0.995	0.990	
4	COIF2	0.985	0.985	0.975	DB10	0.980	0.995	0.980	
5		0.975	1.000	0.975		0.995	0.995	0.990	

Table 4 shows that the size of the window does not influence classification rate by much. However, the overlap factor is playing an important role here. Overall, better results were achieved when r was equal to 2.

In this part, the idea of top to bottom scanning and calculating the three parameters was directly performed on gray scale images to see if high classification rates are obtainable in this case as well. We replaced absolute mean with mean square since pixel values are non-negative. Table 6 represents the average and maximum classification rates using these features which are called as gray tone features in the table. The best average classification rate is 0.944 which is equal to the minimum average rate obtained with wavelet features. Also the maximum obtained rate is 0.975.

Table 6. Average and maximum classification rates with gray tone features

$k \backslash r$	1			2			3		
4	AVR	0.931	0.921	0.944	MAX	0.960	0.970	0.975	
5		0.905	0.906	0.906		0.965	0.970	0.970	

Figure 2 illustrates maximum and averaging over classification rates of each feature set. Comparing the results of gray tone and wavelet features shows that applying wavelet transform to the image before calculating the three parameters has a great effect on having better results. Also, comparing the difference between the maximum and average rates shows that for wavelet features there is a small difference (the biggest difference is for COIF2 and is equal to 0.032) while for gray tone feature this difference is much bigger and is equal to 0.056.

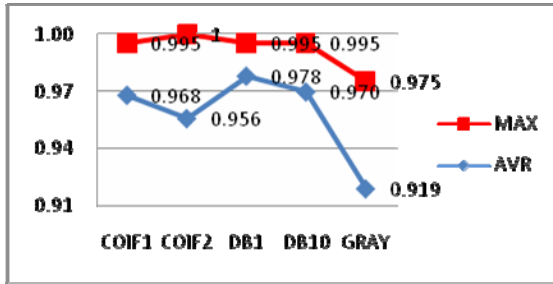


Fig. 2. Classification rate versus feature type. Maximum and average classification of each feature set is shown based on the wavelet filter has been used. The gray tone features are represented with GRAY in the diagram.

The comparison between the results of the proposed methods and others on the same database is represented in Table 7. The best result between those not using HMM is 97% with SVM (Support Vector Machine) and PCA (Principal Component Analysis) coefficients [24]. Among the methods which are proposed based on HMM, the combination of DCT/HMM [27] shows the poorest result. A two dimensional PHHM [26] shows to be more promising. DWT and SHMM [11] reduced the error rate to 3%. Perfect recognitions are achieved in [27] and [9]. However, 9 faces out of

Table 7. Comparative results

Method	Accuracy	Ref
ICA	85	[3]
Eigenface	80.5	[2]
NLPCA	95.5	[22]
Pseudo Zernike	95	[23]
SVM+PCA Coe.	97	[24]
DCT/HMM	84	[25]
2D-PHMM	94.5	[26]
DWT/SHMM	97	[11]
DCT/2D-HMM	100	[27]
Wavelet/HMM	100	[9]
Proposed	100	

10 are used as training set in [27] and the feature extraction method in [9] is much more complicated than our proposed method.

5 Conclusion

In this paper, an efficient human face recognition technique has been presented. The feature extraction is based on wavelet coefficient while the classifier utilized is 1D-CHMM. The proposed method was tested on AT&T face database and high classification rate of up to 100% has been achieved. Comparing the obtained results with those appeared in the literature indicates, the proposed technique requires less computations and easier way of feature extraction, while yielding high accuracy face recognition.

References

1. Kanade, T.: Picture Processing System by Computer Complex and Recognition of Human Faces. Doctoral dissertation, Kyoto University (1973)
2. Turk, M., Pentland, A.: Eigenfaces for Recognition. *J. Cognitive Neuroscience* 3, 71–86 (1991)
3. Yuen, P.C., Lai, J.H.: Face Representation Using Independent Component Analysis. *Pattern Recognition* 35, 1247–1257 (2002)
4. Zhang, J., Yan, Y., Lades, M.: Face Recognition: Eigenface, Elastic Matching and Neural Nets. *Proceedings of the IEEE* 85, 1423–1435 (1997)
5. Rabiner, L.: A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proc. of IEEE* 77, 257–286 (1989)
6. Kohir, V.V., Desai, U.B.: Face Recognition Using DCTHMM Approach. In: *Workshop on Advances in Facial Image Analysis and Recognition Technology (AFIART)*, Freiburg, Germany (1998)
7. Othman, H., Aboulnasr, T.: Hybrid Hidden Markov Model for Face Recognition. In: *IEEE Southwest Symposium on Image Analysis and Interpretation, Texas, USA*, pp. 36–40 (2000)
8. Otsuka, T., Ohya, J.: Recognition of Facial Expressions Using HMM with Continuous Output Probabilities. In: *5th IEEE Workshop on Robot and Human Communication*, Tsukuba, Japan, pp. 323–328 (1996)
9. Bicego, M., Castellani, U., Murino, V.: Using Hidden Markov Models and Wavelets for Face Recognition. In: *12th International Conference on Image Analysis and Processing, USA*, pp. 52–56 (2003)
10. Hung-Son, L., Haibo, L.: Face Identification System Using Single Hidden Markov Model and Single Sample Image per Person. In: *Proceeding of IEEE International Joint Conference on Neural Networks* 1, pp. 455–459 (2004)
11. Nicholl, P., Bouchaffra, D., Amira, A., Perrott, R.H.: Multiresolution Hybrid Approaches for Automated Face Recognition. In: *Second NASA/ESA Conference on AHS*, pp. 89–96 (2007)
12. Bouchaffra, D., Tan, J.: Introduction To Structural Hidden Markov Models: Application to Handwritten Numeral Recognition. *J. Intelligent Data Analysis* 10, 67–79 (2006)
13. Daubechies, I.: *Wavelet Transforms And Orthonormal Wavelet Bases*. Different perspectives on wavelets, San Antonio, 1–33 (1993)

14. Taubman, D., Marcellin, M.W.: JPEG2000 Image Compression: Fundamentals, Standards and Practice. Kluwer, Boston (2002)
15. Grzeczczak, A., Mandal, M.K., Panchanathan, S.: VLSI Implementation of Discrete Wavelet Transform. *IEEE Transaction on VLSI Systems* 4, 421–433 (1996)
16. Lafruit, G., Catthoor, F., Cornelis, J.P.H., De Man, H.J.: An Efficient VLSI Architecture for 2-D Wavelet Image Coding with Novel Image Scan. *IEEE Transaction on VLSI Systems* 7, 56–68 (1999)
17. McCanny, P., Masud, S., McCanny, J.: Design and Implementation of the Symmetrically Extended 2D Wavelet Transform. *Proceeding of ICASSP* 3, 3108–3111 (2002)
18. Shahbahrami, A., Juurlink, B., Vassiliadis, S.: Implementing the 2-D Wavelet Transform on SIMD-Enhanced General-Purpose Processors. *IEEE Transaction on Multimedia* 10, 43–51 (2008)
19. Mallat, S.: A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 11, 674–693 (1989)
20. Chen, M.Y., Kundu, A., Srihari, S.N.: Variable Duration Hidden Markov and Morphological Segmentation for Handwritten Word Recognition. *IEEE Transaction on Image Processing* 4, 1675–1688 (1995)
21. Rabiner, L.R., Jung, B.-H.: *Fundamentals of Speech Recognition*. Prentice Hall, Englewood Cliffs (1993)
22. Makaremi, I., Araabi, B.N.: Face Recognition Using Neural Network Based Nonlinear Principal Component Analysis. In: *Proceedings of the 8th International Conference on Pattern Recognition and Information Processing*, Minsk, Belarus (2005)
23. Nabatchian, A., Abel-Raheem, E., Ahmadi, M.: Human Face Recognition Using Different Moment Invariants: A Comparative Study. In: *The 2008 International Congress on Image and Signal Processing (CISP 2008)*, Sanya, Hainan, China (2008)
24. Guo, G., Li, S.Z., Kapluk, C.: Face Recognition By Support Vector Machines. *Image and Vision Computing* 19, 631–638 (2001)
25. Nefian, A., Hayes, M.: Hidden Markov Models for Face Recognition. In: *ICASSP 1998*, pp. 2721–2724 (1998)
26. Samaria, F.: *Face Recognition Using Hidden Markov Models*, Ph.D. dissertation. Cambridge University Engineering Department (1994)
27. Othman, H., Aboulnasr, T.: A Separable Low Complexity 2D HMM with Application to Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25, 229–238 (2003)

Face Recognition Based on Histogram of Modular Gabor Feature and Support Vector Machines

Xiaodong Li, Shumin Fei, and Tao Zhang

Key Laboratory of Measurement and Control of Complex Systems of Engineering, Ministry of Education, Southeast University, Nanjing 210096, China
School of Automation, Southeast University, Nanjing 210096, China

Abstract. In this paper, a novel face recognition algorithm based on histogram of modular Gabor feature and support vector machines is proposed. In this method, each face image is separate into several parts on which Gabor transformation is performed, respectively and then employed 2DPCA for dimensionality reduction. Subsequently, histogram sequences are calculated based on these coefficient features. The final features of face image can be obtained by the fusion of the normalized histogram sequences using weight scheme. Finally, support vector machines is used as classifier. Several experiments on popular face databases such as CAL-PEAL and FERET demonstrate the effectiveness of the proposed method.

Keywords: face recognition, Gabor wavelet, support vector machines (SVM), histogram sequence.

1 Introduction

Face recognition has been an important issue in image processing and pattern recognition over the last several decades. It plays an important role in many applications such as card identification, access control, mug shot searching and security monitoring. So it has become a significant research field [1].

To obtain best recognition performance, feature extraction and classification are two important problems that should be paid attention to. As to feature extraction, much progress has been made under controlled conditions as described in [1-3]. Eigenfaces [4] method introduced by Turk and Pentland and Linear Discriminant Analysis (LDA)[5],[6] are two popular approaches used in face recognition. As we all know, most existing face recognition method are all based on original gray image, and it is difficult to get local information which is important to face recognition. In recent years, many methods based on Gabor filters have been proposed [7-10]. Because the Gabor filters exhibit desirable characteristics of spatial localization and orientation selectivity, and the Gabor filter representations of face image (termed also as Gabor-faces) are robust to illumination and expressional variability, they are used extensively in face recognition. However, the dimensionality of the Gabor feature space is overwhelmingly high, because the Gaborfaces are obtained by convolution of the face image with dozens of Gabor filters. Therefore, many sampling or compressing methods are proposed to reduce the space dimension to avoid dealing with the enormous

datum [11],[12],[13]. Down-sample method was used in [11], but losing some discriminant information is its drawback. In [12], predetermined fiducial points at face landmarks of each face was selected before performing Gabor transformation. However, it is difficult to locate the eyes. [13] applied the Gabor feature histogram sequence to do dimension reduction and got a good recognition performance. In this paper, modular Gabor transformation and histogram sequence scheme are used to extract discriminant feature to improve the recognition performance.

When considering classification, there are several conventional methods, for example, k-nearest neighborhood classification, template matching and so on. In practical face recognition problems, they are inefficient and time-consuming. In recent years, neural networks[14] are used to solve the above problem. However, neural networks have some internal shortcomings, such as difficulty for determining the structure of neural networks, over-learning and getting into local extremes easily. To address these problems, support vector machines developed principally by Vapnik[15], have drawn much attention and been applied successfully in recent years.

According to the analysis mentioned above, a novel face recognition algorithm based on histogram of modular Gabor feature(HMG) and support vector machines is proposed. Fig.1 represents the idea of the proposed algorithm. In this method, each face image is separated into several parts, on which Gabor transformation and 2DPCA are performed in turn, respectively. Subsequently, histogram sequences are calculated based on these coefficient features. At the same time, an algorithm of weight calculation is presented. The final features of a face image can be obtained by weight fusion of the histogram sequence corresponding to each modular. At last, support vector machines are used as classifiers. Several experiments on popular face databases such as CAL-PEAL and FERET demonstrate the effectiveness of the proposed method.

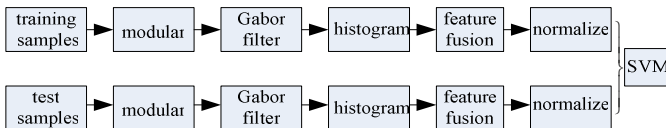


Fig. 1. The framework of the proposed method

The paper of the rest is organized as follows: Section 2 introduces the feature extraction method. The concept of support vector machines is given in Section 3. The last two sections give experiment results and conclusions.

2 Feature Extraction

In this section, the method of feature extraction is proposed in detail.

2.1 Gabor Wavelets

Gabor wavelets were introduced to image analysis due to their biological relevance and computational properties. The Gabor wavelets, whose kernels are similar to the 2-D receptive field profiles of the mammalian cortical simple cells, exhibit desirable

characteristics of spatial locality and orientation selectivity, and are optimally localized in the space and frequency domains. The Gabor wavelets (kernels, filters) can be defined as follows [16]:

$$\psi_{\mu,\nu}(z) = \frac{\|k_{\mu,\nu}\|^2}{\sigma^2} e^{\left(-\|k_{\mu,\nu}\|^2 \|z\|^2 / 2\sigma^2\right)} \left[e^{ik_{\mu,\nu}z} - e^{-\sigma^2 / 2} \right]. \tag{1}$$

where μ and ν denotes orientation and scale of the Gabor kernels, respectively. $z = (x, y)$. $\|\bullet\|$ denotes the norm operator, and the wave vector $k_{\mu,\nu}$ is defined as follows:

$$k_{\mu,\nu} = k_\nu e^{i\phi_\mu} \tag{2}$$

where $k_\nu = k_{\max} / f^\nu$ and $\phi_\mu = \pi\mu / 8$. k_{\max} is the maximum frequency, and f is the spacing factor between kernels in the frequency domain.

The Gabor kernels in (1) are all self-similar since they can be generated from one filter, the mother wavelet, by scaling and rotation via the wave vector $k_{\mu,\nu}$. Each kernel is a product of a Gaussian envelope and a complex plane wave, while the first term in the square brackets in (1) determines the oscillatory part of the kernel and the second term compensates for the DC value. The effect of the DC term becomes negligible when the parameter, which determines the ratio of the Gaussian window width to wavelength, has sufficiently large values.

The Gabor wavelet representation of an image is the convolution of the image with a family of Gabor kernels as defined by (1). Let $\mathbf{I}(x, y)$ be the gray level distribution of an image, the convolution of image and a Gabor kernel $\psi_{\mu,\nu}$ is defined as follows:

$$\mathbf{O}_{\mu,\nu}(z) = \mathbf{I}(z) * \psi_{\mu,\nu}(z). \tag{3}$$

where $z = (x, y)$, $*$ denotes the convolution operator, and $\mathbf{O}_{\mu,\nu}(z)$ is the convolution result corresponding to the Gabor kernel at orientation μ and ν scale. In most cases, one would use Gabor wavelets of five different scales, $\nu \in \{0..4\}$, and eight orientations, $\mu \in \{0..7\}$. Therefore, the set $S = \{\mathbf{O}_{\mu\nu}(z) : \mu \in \{0..7\}, \nu \in \{0..4\}\}$ forms the Gabor wavelet representation of the image $\mathbf{I}(z)$.

To encompass different spatial frequencies (scales), spatial localities, and orientation selectivity, we concatenate all these representation results and derive an augmented feature vector \mathbf{X} .

$$\mathbf{X} = (\mathbf{O}_{0,0} \mathbf{O}_{0,1} \cdots \mathbf{O}_{4,7}). \tag{4}$$

2.2 Modular Gabor Feature Extraction

Each part corresponding to eyes, nose and mouth have more important discriminant feature than other parts of a face image. In this paper, a modular scheme is applied to each face image, and Fig. 2 illustrates our approach.

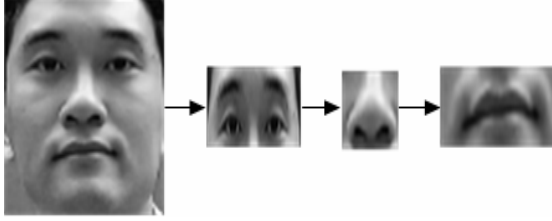


Fig. 2. The modular scheme

Suppose there are c known pattern classes, each class contains n samples. $\mathbf{I}_{ij}(z)$ denotes the i th class j th face image, $i = 1, 2, \dots, c$, $j = 1, 2, \dots, n$, so $\mathbf{I}_{ijk}(z)$, $k \in \{1, 2, 3\}$ is used to indicate the k th part of the i th class j th face image. According to Eq.(3) and (4), the Gabor features of $\mathbf{I}_{ijk}(z)$ are \mathbf{X}_{ijk} .

Because Gabor feature space is overwhelmingly high, dimensionality reduction algorithm should be employed to \mathbf{X}_{ijk} . 2DPCA is a good choice for its simplicity and powerful represent. Suppose \mathbf{A}_k ($k \in \{1, 2, 3\}$) denotes the transformation matrix corresponding to k th modular in the face image, and the feature in the new space can be obtained by following linear transformation:

$$\mathbf{Y}_{ijk} = \mathbf{A}_k^T \mathbf{X}_{ijk} . \quad (5)$$

2.3 Gabor Histogram Feature Extraction

In order to further reduce the dimension of Gabor feature, histogram operation is used to statistic the Gabor feature. The histogram of image $\mathbf{I}(x, y)$ whose gray level range is $[0, L-1]$ is defined as follows[17]:

$$h_l = \sum_{x,y} \phi\{\mathbf{I}(x, y) = l\}, l = 0, 1, \dots, L-1 . \quad (6)$$

where l is the l th gray level, h_l is the number of pixels in l th gray level.

Therefore, the Gabor histogram feature of k th modular are:

$$\mathbf{H}_{ijk} = (h_{ijk}^1, h_{ijk}^2, \dots, h_{ijk}^{L-1}) . \quad (7)$$

where $h_{ijk}^l = \sum \phi\{Y_{ijk} = l\}, l = 0, 1, \dots, L-1 . i = 1, 2, \dots, c, j = 1, 2, \dots, n, k \in \{1, 2, 3\}$.

As we all know, histogram feature can not be classified directly using traditional classification, so normalization operation is necessary for a convenient representation and comparison. We call histogram similarity measure as HSM.

2.4 Fusion Feature Based on Weight Scheme

At present, a popular modular feature fusion method is to concatenate all features of each modular together. However, this algorithm dose not take into account the fact

that different part plays different role, so giving different weight to different part is necessary when fusion these features. The method of calculating the weight is described as follows:

- (1) Calculate the average of within-class sample distance corresponding to each modular $L_w(k), k \in \{1,2,3\}$;
- (2) Calculate the average of between-class sample distance corresponding to each modular $L_b(k), k \in \{1,2,3\}$;
- (3) Calculate $L(k) = L_b(k)/L_w(k)$, it is easy to see that the bigger the $L(k)$ is, the role of the k th part plays is important.
- (4) Suppose $w(k), k \in \{1,2,3\}$ is the weight corresponding to each modular of face image, the weight can be determined using the following formulary:

$$w(k) = L(k) / \sum_k L(k). \quad (8)$$

Based on the analysis, the final features of a face image are as follows:

$$\mathbf{H}_{ij} = \sum_{k=1}^3 \mathbf{H}_{ijk} w(k), i = 1,2,\dots,c, j = 1,2,\dots,n, k \in \{1,2,3\}. \quad (9)$$

In order to make the features convenient to be used in several kinds of classifiers, the Gabor histogram feature \mathbf{H}_{ij} should be normalized.

3 Support Vector Machines [15], [18]

The basic idea of SVM is to transform the signal to a higher dimensional feature space and find the optimal hyper-plane in the space that maximizes the margin between the classed.

SVM stems from statistical learning theory. It minimize a bound on the empirical error and the complexity of the classifier at the same time. Accordingly, It is capable of learning in spare high-dimensional spaces with relatively few training examples. Let $\{\mathbf{x}_i, y_i\}, i = 1,2,\dots,N$ denotes N training examples, in which x_i comprises an M -dimensional pattern and y_i is its class label. Without any loss of generality, we shall confine ourselves to the two class pattern recognition problem. That is to say, $y_i \in \{+1,-1\}$. We agree that $y_i = +1$ is assigned to positive examples, whereas $y_i = -1$ is assigned to counter examples.

The data to be classified by the SVM might be linearly separable in their original domain or not. If they are separable, then a simple linear SVM can be used for their classification. However, the power of SVM is demonstrated better in the nonseparable case, when the data cannot be separated by a hyperplane in their original domain. In the latter case, we can project the data into a higher dimensional Hilbert space and attempt to linearly separate them in the higher dimensional space using kernel functions. Let Φ denote a nonlinear map $\Phi: \mathcal{X}^M \rightarrow \mathbf{H}$, where \mathbf{H} is a higher dimensional

Hibert space. SVMs construct the optimal separating hyperplane in H . Therefore, their decision boundary is of the form:

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^N a_i y_i K(\mathbf{x}, \mathbf{x}_i) + b\right). \tag{10}$$

where $K(\mathbf{x}, \mathbf{x}_i)$ is a kernel function that defines the dot product between $\Phi(\mathbf{x})$ and $\Phi(\mathbf{x}_i)$ in H , and a_i are the nonnegative Lagrange multipliers associated with the quadratic optimization problem that aims to maximize the distance between the two classes measured in \mathbf{H} subject to the constraints:

$$\mathbf{w}^T \Phi(\mathbf{x}) + b \geq 1 \text{ for } y_i = +1. \tag{11}$$

$$\mathbf{w}^T \Phi(\mathbf{x}) + b \leq -1 \text{ for } y_i = -1. \tag{12}$$

Kernel function plays an important role. Frequently used kernel functions are polynomial kernel function and radial basis function Gaussian kernel function, which are respectively

$$K(\mathbf{x}_i, \mathbf{x}_j) = (m\mathbf{x}_i^T \mathbf{x}_j + n)^n. \tag{13}$$

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right). \tag{14}$$

Most of cases in practical are multi-classed, thus, we have to design an approach to extend the application of SVM to a multi-classifying field for which the SVM can deal with only two classes. The different combination principles constitute different classifying algorithm. In this paper, the Libsvm tool box[19] which can deal with multi-classifying is used as classifier to simulate the proposed method.

4 Experiment Results

In this section, several experiments were designed to demonstrate the effectiveness of our proposed method. In order to show the recognition performance in an all-round way, we compare the proposed method with other popular feature extraction methods such as PCA, LDA, Gabor transformation, Modular Gabor(MG)and histogram of modular Gabor Feature(HMG). At the same time, SVM classifier used in this paper is compared with minimal distance classifier with Euclidean distance (MD) and BP neural networks(BPNN). The first experiment is conducted on a subset of CAS-PEAL database, and the second one is conducted on the FERET database.

4.1 Experiment Using the CAS-PEAL Database

The CAS-PEAL face database [20],[21] contains 30,900 images of 1040 individuals with varying Pose, Expression, Accessory, and Lighting (PEAL). we select the frontal images from the subsets of accessory, distance, background, expression, lighting and aging as probe sets. Arbitrary 20 face images of each person from 100 individuals are

used in our experiments, therefore there are 2000 pictures in total. The face portion of each image is manually cropped and then normalized to 50×60 pixels. Some samples of this database are showed in Fig.3.



Fig. 3. Sample images of some persons in the CAS-PEAL database

In this experiment, the training and testing set are selected randomly for each individual. The number of training samples of each person is set 5,7 and 9, respectively, then the corresponding remaining samples are used for test. We repeat the recognition procedure 10 times by choosing different training and testing sets, respectively. At last, the minimal distance classifier with Euclidean distance is employed for classification. The optimal average recognition rates corresponding to each method versus the number of training samples are illustrated in Tab.1.

Table 1. Optimal recognition rate(%)of each method with different number of training samples

Methods	5	7	9
PCA+MD	34.71	38.88	40.26
LDA+MD	42.35	45.21	46.30
Gabor+MD	50.12	55.46	58.79
MG+MD	61.56	64.21	66.53
HMG+HSM	63.22	65.39	68.28
HMG+MD	63.78	66.86	69.49
HMG+BPNN	65.92	67.22	71.31
HMG+SVM	68.48	71.68	75.63

Tab.1 shows that the Gabor features are better than popular feature such as PCA feature and LDA feature in improvement of recognition performance. we also can see that the histogram similarity measure(HSM)is not the best classification method for histogram feature, and histogram normalization plus traditional classifier can improve recognition performance. As indicating in this table, SVM is a better classification than MD and NN.

4.2 Experiment Using the FERET Database

The proposed method was also tested on a subset of the FERET database. The FERET face image database is a result of the FERET program, which was sponsored by the

US Department of Defense through the DARPA Program[22], [23]. It has become a standard database for testing face recognition algorithm. This subset includes 1400 images of 200 persons (each person has seven images), which involve variations in facial expression, illumination, and pose. In our experiment, the facial portion of each original image is cropped manually based on the location of eyes and resized to 40×40 pixels without histogram equalization. Some facial portion images of one person are shown in Fig.4.



Fig. 4. Sample images of some persons in the FERET database

In our experiments, random 5 images of each individual are used for training, and the remaining 2 images are used for test. We repeat the procedure 10 times, and the average result is used as the final recognition rate. PCA, LDA and other method corresponding to Gabor transformation are, respectively, used for feature extraction. Finally, MD, NN and SVM are employed for classification. The recognition rate versus feature extraction method and classifiers are plotted in Tab.2.

Table 2. Comparison of recognition rate(%) corresponding to different method

Methods	PCA	LDA	HMG+HSM	HMG+MD	HMG+BPNN	HMG+SVM
Accuracy	50.38	53.13	74.33	74.64	77.39	80.28

Tab.2 indicates that in the FERET face database, the proposed method in this paper is effective to increase the recognition rate.

5 Conclusions

Gabor feature is robust to the variation of illumination, expression, pose, and so on, while support vector machines has many virtues in dealing with classification. Therefore, the fusion of them is a good scheme for face recognition. To further improve the performance of face recognition using Gabor feature and support vector machines, a method based on Histogram of Modular Gabor Feature and Support Vector Machines is proposed in this paper. There are two novel aspects in this algorithm. On the one hand, histogram of modular Gabor feature is used for further dimensionality reduction; on the other hand, weight trick is applied to fusion the modular histogram feature. The results of several experiments demonstrate the effectiveness of the proposed method.

Acknowledgment

This work is support by the National Science Foundation of China under Grant No.60835001.

References

1. Tan, X.Y., Chen, S.C.: Face Recognition from a Single Image per Person: a Survey. *Pattern Recognition* 391, 1725–1745 (2006)
2. Phillips, P.J., Flynn, P.J., Scruggs, T., et al.: Overview of the Face Recognition Grand Challenge. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1, 947–954 (2005)
3. Zhao, W., Chellappa, R., Rosenfeld, A., Phillips, P.J.: Face Recognition: a Literature Survey. *Computing Surveys* 35(4), 399–458 (2003)
4. Turk, M., Pentland, A.: Eigenfaces for Recognition. *Journal of Cognitive Neuroscience* 3(1), 71–86 (1991)
5. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition using Class Specific Linear Projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(7), 711–720 (1997)
6. Zhi, R.C., Ruan, Q.Q.: Two-dimensional Direct and Weighted Linear Discriminant Analysis for Face Recognition. *Neurocomputing* 71(16-18), 3607–3611 (2008)
7. Choi, W.P., Tse, S.H., Wong, K.W., Lam, K.M.: Simplified Gabor Wavelets for Human Face Recognition. *Pattern Recognition* 42(3), 1186–1199 (2008)
8. Shen, L.L., Li, B., Fairhurst, M.: General Discriminant Analysis for Face Identification and Verification. *Image and Vision Computing* 25(5), 553–563 (2007)
9. Wang, L., Li, Y.P.: A Novel 2D Gabor Wavelets Window Method for Face Recognition. In: Günsel, B., Jain, A.K., Tekalp, A.M., Sankur, B. (eds.) *MRCSS 2006*. LNCS, vol. 4105, pp. 497–504. Springer, Heidelberg (2006)
10. Loris, N., Dario, M.: Weighted Sub-Gabor for Face Recognition. *Pattern Recognition Letters* 28(4), 487–492 (2007)
11. Wang, L., Li, Y.P., Wang, C.B., Zhang, H.Z.: 2D Gabor Face Representation Method for Face Recognition with Ensemble and Multichannel Model. *Image and Vision Computing* 26, 820–828 (2008)
12. Pan, X., Ruan, Q.Q.: Palmprint Recognition using Gabor Feature-based (2D)2PCA. *Neuro Computing* 71(13-15), 3032–3036 (2008)
13. Zhang, W.C., Shan, S.G., Zhang, H.M., Chen, J., Chen, X.L., Gao, W.: Histogram Sequence of Local Gabor Binary Pattern for Face Description and Identification. *Journal of Software* 17(12), 2508–2517 (2006)
14. Jing, X.Y., Yao, Y.F., Yang, J.Y., Zhang, D.: A Novel Face Recognition Approach based on Kernel Discriminative Common Vectors (KDCV) Feature Extraction and RBF Neural Network. *Neurocomputing* 71(13-15), 3044–3048 (2008)
15. Vapnik, V.: *The Nature of Statistical Learning Theory*. Springer, New York (1995)
16. Liu, C.J., Wechsler, H.: Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition. *IEEE Transactions on Image processing* 11(4), 467–476 (2002)
17. Zhang, B.C., Shan, S.G., Chen, X.L., Gao, W.: Histogram of Gabor Phase Patterns (HGPP): A Novel Object Representation Approach for Face Recognition. *IEEE Transactions on Image Processing* 16(1), 57–68 (2007)
18. Zhang, X.G.: *Introduction to Statistical Learning Theory and Support Vector Machines*. *Acta Automatica Sinica* 26(1), 32–42 (2000)

19. LIBSVM: A Library for Support Vector Machines,
<http://www.csie.ntu.edu.tw/~cjlin/libsvm>
20. Gao, W., Cao, B., Shan, S.G., Zhou, D.L., Zhang, X.H., Zhao, D.B.: The CAS-PEAL large scale Chinese face database and evaluation protocols. Technical Report, No. JDL_TR_04_FR_001, Joint Research & Development Laboratory, CAS (2004)
21. Gao, W., Cao, B., Shan, S.G., Chen, X.L., Zhou, D.L., Zhang, X.H., Zhao, D.B.: The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations. *IEEE Transactions on System Man, and Cybernetics (Part A)* 38(1), 149–161 (2004)
22. Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J.: The FERET Evaluation Methodology for Face-Recognition algorithms. *IEEE Transactions. on Pattern Analysis and Machine Intelligence* 22(10), 1090–1104 (2000)
23. Phillips, P.J.: The Facial Recognition Technology (FERET) database (2004),
http://www.itl.nist.gov/iad/humanid/feret/feret_master.html

Feature-Level Fusion of Iris and Face for Personal Identification

Zhifang Wang¹, Qi Han¹, Xiamu Niu¹, and Christoph Busch²

¹ School of Computer Science and Technology, Harbin Institute of Technology
Heilongjiang, China

² Norwegian Information Security laboratory, *Gjøvik University College*
Gjøvik, Norway

{zhifang.wang, qi.han, xiamu.niu}@ict.hit.edu.cn
christoph.busch@hig.no

Abstract. Feature-level fusion remains a challenging problem for multimodal biometrics. However, existing fusion schemes such as sum rule and weighted sum rule are inefficient in complicated condition. In this paper, we propose an efficient feature-level fusion algorithm for iris and face in parallel. The algorithm first normalizes the original features of iris and face using z-score model, and then take complex FDA as the classifier of unitary space. The proposed algorithm is tested using CASIA iris database and two face databases (ORL database and Yale database). Experimental results show the effectiveness of the proposed algorithm.

Keywords: Biometrics, Feature-level, Parallel fusion, Unitary space, CFDA.

1 Introduction

Biometrics refers to the automatic personal identification by using something that you are (e.g. iris or face) or something that you do or produce (e.g. voice or handwriting signature) [1]. However, the performances of unimodal biometric systems have to contend with a variety of problems such as background noise, signal noise and distortion, and environment or device variations [2]. Therefore, multimodal biometric systems are proposed to solve the above mentioned problems of unimodal biometrics systems. So, many literatures and algorithms are presented to do the research about multimodal biometric fusion [3,4,5]. Along with the fusion in match score level [6,7] and decision level [8,9], one important branch is to do the fusion in feature level [10-12]. Among the existing research works, sum rule and weighted sum rule are the popular fusion scheme [3,4]. However, they are still inefficient in complicated application environment. Hence, we are motivated to design a feature-level fusion algorithm that could be more reliable and accurate.

In this paper, we proposed a novel fusion algorithm for iris and face in feature level based on complex vector because face recognition is friendly and non-invasive whereas iris recognition is one of the most accurate biometrics. This algorithm normalizes the original features of iris and face using z-score model before fusing them. Then the two normalized feature sets of iris and face are fused as the complex vectors

and form unitary space. Finally, we take the complex Fisher discriminate analysis (CFDA) [11] as a classifier of unitary space. The contributions of this paper are as follows: (1) Z-score normalization model is adopted to eliminate the difference of the order of magnitude and the distribution of the features derived from iris and face, which makes the system more accurately; (2) Compared with sum rule and weighted sum rule, the experiments shows the effectiveness of the proposed fusion scheme; (3) CFDA is used to resolve the classification problem of the unitary space and ensure the availability of the proposed scheme.

The rest of this paper is organized as follows: In the next section, we present some preliminaries: feature extraction and normalization. Section 3 focuses on the fusion scheme and classification. In section 4, experimental results and comparison are given. Finally, we conclude in section 5.

2 Preliminaries

In this section, we first present the framework of the multimodal biometric algorithm as Fig. 1. The algorithm comprises 3 phases. In the first phase, the features of iris and face are extracted respectively. We then normalize the features before fusion. Finally, we fuse the normalized features in parallel and use CFDA to classify. The following content will describe the preliminaries: feature extraction and normalization.

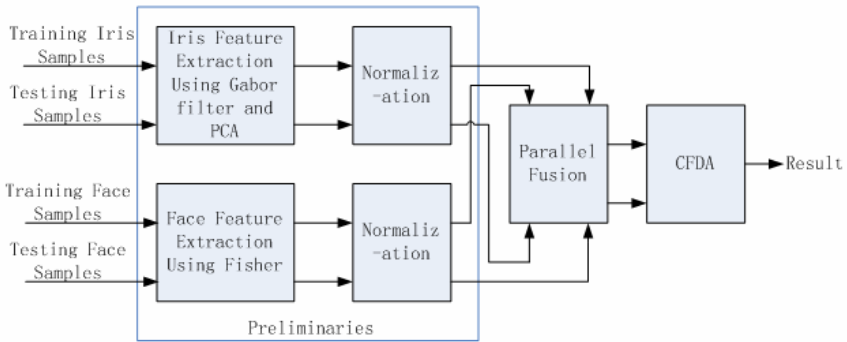


Fig. 1. The framework of the proposed algorithm

2.1 Feature Extraction

In this algorithm, the objects for fusion are iris features and face features. So the first step is to obtain them. This section describes the feature extraction of iris and face respectively. For face recognition, principle component analysis (PCA) [13] and Fisher discriminant analysis (FDA) [14] are two notable methods. In order to extract more discriminated feature, we adopt the latter method because the performance of FDA is better than PCA.

For iris recognition, Gabor filter is the popular feature extractor. For example, Daugman [15] takes 2D Gabor filter while Tan [16] uses 2D even Gabor filter. However, the feature attained by Daugman’s method is not convenient for fusion with face feature,

which is represented as binary vector. Therefore, this algorithm adopts the latter method and obtains the iris feature represented as the real vector. Whereas the dimension of iris features derived by Tan’s method is too large, PCA are used to solve this problem and control the dimension of iris feature equal to that of face feature.

2.2 Normalization

Traditionally, feature-level fusion methods directly fuse two kinds of features after feature extraction. As we know, due to the difference of the modal and extraction method, the order of magnitude and the distribution between iris feature and face feature might be different. In order to eliminate the unbalance and get good performance, we are motivated to normalize the feature before fusion using z-score model.

Let a_j^i be a d -dimension iris feature of the j th iris training sample from the i th class, and b_j^i denotes a d -dimension face feature of the j th face training sample from the i th class. Then the iris feature set and the face feature set are respectively represented as $A = (a_1^1, \dots, a_m^1, a_1^2, \dots, a_m^n)$ and $B = (b_1^1, \dots, b_m^1, b_1^2, \dots, b_m^n)$.

Let A_k is the k th row of the iris feature set A . We use the following method to get the corresponding normalized component X_k . Firstly, compute

$$C_k = \frac{A_k - \overline{A_k}}{\sigma_k}, \tag{1}$$

where $\overline{A_k}$ denotes the mean value of A_k , and σ_k is the standard deviation of A_k . Then, we can get the normalized component by

$$X_k = \frac{C_k - C_{\min}}{C_{\max}}, \tag{2}$$

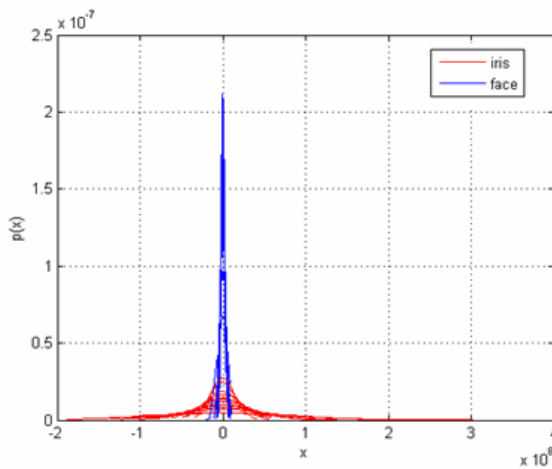


Fig. 2. The distribution of the original components

where C_{\min} and C_{\max} denote the minimum value and the maximum value of C_k respectively. The normalized feature set is $X = (X_1 \cdots X_d)^T$. For face feature, repeat the same procedure and get the normalized feature set $Y = (Y_1 \cdots Y_d)^T$. Fig. 2 gives the distribution of the original components. After the normalization, the distribution is changed as fig. 3. From them, it can be found that the order of magnitude and the distribution of two kinds of features are similar after normalization.

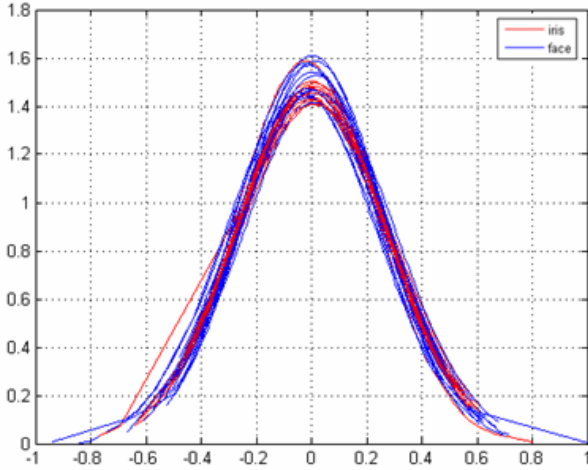


Fig. 3. The distribution of the normalized components

3 Fusion and Classification

In this section, we first present the parallel fusion [12] for the normalized feature. Then CFDA [13] is used as the classifier to classify the fusion features.

Let $x = (x_1 \cdots x_d)$ denotes a normalized iris feature vector, $y = (y_1 \cdots y_d)$ is a normalized face feature vector. The fusion feature ξ in sum rule can be defined as $\xi = (x_1 + y_1, \dots, x_d + y_d)$. For weighted sum rule, we take $\theta = 3/7$ as the weighted parameter because the performance of iris recognition is better than that of face recognition. The fusion feature in weighted sum rule can be denoted as $\xi = (x_1 + \theta y_1, \dots, x_d + \theta y_d)$. So, sum rule can also be considered as the special case of weighted sum rule. In this paper, we adopt a novel fusion method: parallel fusion. The format of the fusion feature is defined as $\xi = (x_1 + i y_1, \dots, x_d + i y_d)$ (i is the imaginary unit). This method combines two kinds of feature into a complex vector and considers the classification in unitary space.

In unitary space, the within-class scatter matrix, the between-class matrix and the total scatter matrix are respectively defined as follows:

$$S_w = \sum_{i=1}^n P(\omega_i) E\{(\xi - \bar{\xi}_i)(\xi - \bar{\xi}_i)^H \mid \omega_i\}, \tag{3}$$

$$S_b = \sum_{i=1}^n P(\omega_i) E\{(\bar{\xi}_i - \bar{\xi})(\bar{\xi}_i - \bar{\xi})^H\}, \tag{4}$$

$$S_t = S_w + S_b = E\{(\xi - \bar{\xi})(\xi - \bar{\xi})^H\}, \tag{5}$$

where $P(\omega_i)$ is the prior probability of class i ; $\bar{\xi}_i$ is the mean vector of the features in class i ; $\bar{\xi}$ is the mean vector of all the features; H is the denotation of conjugate transpose.

The criterion function of CFDA can be defined as follows:

$$J_{\text{cfda}}(\phi) = \frac{\phi^H S_b \phi}{\phi^H S_w \phi} \tag{6}$$

Our goal tries to find the optimal projection vector ϕ which maximum the criterion. The solution of this problem is presented as follows:

- Step 1: compute the eigenvalues $\lambda_1, \dots, \lambda_d$ and the corresponding eigenvectors v_1, \dots, v_d of the within-class scatter matrix S_w , and then attain the transformation matrix $W = V\Delta^{-1/2}$ where $V = (v_1, \dots, v_d)$ and $\Delta = (\lambda_1, \dots, \lambda_d)$
- Step 2: let $\tilde{S}_b = W^H S_b W$ and compute its orthonormal eigenvectors η_1, \dots, η_p corresponding to p largest eigenvalues. Then the optimal projection vectors are $\phi_1 = W\eta_1, \dots, \phi_k = W\eta_p$

Then we use the optimal projection vector to extract the projection feature of the fusion feature. Let $P = (\phi_1, \dots, \phi_p)$, and ξ is a fusion feature, we can obtain the projection feature z by the following transformation:

$$z = P^H \xi. \tag{7}$$

In unitary space, the measurement can be defined by:

$$\|z\| = \sqrt{z^H z}. \tag{8}$$

Correspondingly, the distance in unitary space between the complex vectors z_1 and z_2 is defined as follows:

$$\|z_1 - z_2\| = \sqrt{(z_1 - z_2)^H (z_1 - z_2)}. \quad (9)$$

Finally, based on this distance, we can classify the features.

4 Experiments

The experiments are performed on CASIA iris image database (ver. 1.0) [17] and two face databases (ORL database [18] and Yale database [19]). We take two experiments: experiment I fuses CASIA iris features with ORL face features; experiment II fuses CASIA iris features with Yale face features. Our goal is to compare our algorithm with two unimodal biometrics (iris [16] and face [14]), and other two fusion approaches (sum rule and weighted sum rule) using same features as our algorithm.

4.1 Database

CASIA iris image database (ver. 1.0) includes 756 iris images from 108 eyes (hence 108 classes). For each eye, 7 images are captured in two sessions, where three samples are collected in the first session and four in the second session. Three samples of the first session are taken as the training samples. Other samples are used to test the performance of the algorithm.

Two face databases are used in this paper. One is ORL face database which includes 40 people, 10 different images with pose and expression variation per person. The other is Yale face database. It contains 11 images of 15 people in a variety of conditions including with and without glasses, illumination variation, and changes in facial expression.

In order to fuse the different features derived from iris and face, the dimension and the number of the feature should be equal. The former problem has been solved in the process of the feature extraction. Aim to the latter, this paper adopts the following rule: according to the number of the sample in the face database, we randomly select the same number of the iris sample. For example, because ORL face database includes 40 people, we choose 40 eyes at random to fuse. And the previous 7 images per person of ORL database are used. 3 images per person are selected as the training samples, the remainder images are taken as the testing set.

4.2 Experimental Results

False reject rate (FRR) and false accept rate (FAR) are usually used to test the performance of the system. However, False match rate(FMR) and false non-match rate(FNMR) are more suitable to evaluate the performance of the algorithms in an off-line technology test like this as failure to enroll rate(FTE) and failure to acquire rate(FTA) are not available[20]. Thus, FMR and FNMR are used as the performance parameters of the proposed algorithm in this paper. Equal error rate (EER) is also taken as a performance parameter.

Table 1. The comparison results of the performance (EER %)

Algorithm	ORL	Yale
Iris	3.11	3.33
Face	3.75	7.46
Sum rule	2.37	6.67
Weighted sum rule	3.05	8.00
Proposed algorithm	0.07	2.9

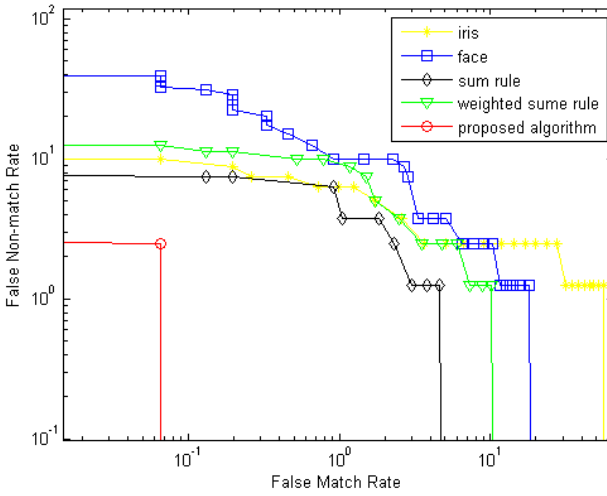


Fig. 4. DET curve of experiment I

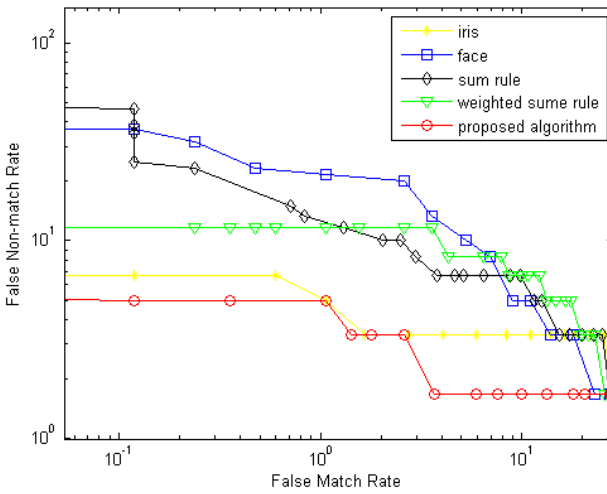


Fig. 5. DET curve of experiment II

Table 1 shows the comparison results of EER between the proposed algorithm and other four algorithms by two experiments. From table 1, it is evident that EER of unimodal biometrics is efficiently reduced after parallel fusion in proposed method. Comparing with sum rule and weighted sum rule, our algorithm is still better. In order to present the whole performance of proposed algorithm, we give the DET curves of two experiments as fig. 4-5. From fig. 4-5, it can be obviously found that our algorithm has much efficiency than other algorithms both in experiment I and experiment II.

5 Conclusion

In this paper, a feature-level fusion algorithm of iris and face is proposed for personal identification. The algorithm uses z-score normalization model to eliminate the difference of the order of magnitude and the distribution between iris features and face features. Then we fuse the normalized features in parallel and take CFDA as a classifier of unitary space. We have experimented on CASIA iris database and two face database (ORL database and Yale database). Experiments show that our algorithm improves the performance of two unimodal biometrics and outperforms sum rule fusion and weighted sum rule fusion.

Acknowledgements. The author would like to thank Chinese Academy of Sciences for sharing their database of iris images. This work is supported by the National Natural Science Foundation of China (Project Number: 60832010,60671064,60703011), the Chinese national 863 Program (Project Number:2007AA01Z458) and the Research Fund for the Doctoral Program of Higher Education (RFDP: 20070213047).

References

1. Ortega-Garcia, J., Bigun, J., Reynolds, D., Gonzalez-Rodriguez, J.: Authentication gets Personal with Biometrics. *Signal Processing Magazine* 21, 50–62 (2004)
2. Khan, M.K., Zhang, J.S.: Multimodal Face and Fingerprint Biometrics Authentication on Space-limited Token. *Neurocomputing* 71, 3026–3031 (2008)
3. Kittler, J., Hatef, M., Duin, R.P.W., Matas, J.: On Combining Classifiers. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 20, 226–239 (1998)
4. Ross, A., Jain, A.: Information Fusion in Biometrics. *Pattern Recognition Letters* 24, 2115–2125 (2003)
5. Jain, A.K., Ross, A.: Multibiometric Systems. *Communications of the ACM* 47, 34–40 (2004)
6. Jain, A., Nandakumar, K., Ross, A.: Score Normalization in Multimodal Biometric Systems. *Pattern Recognition* 38, 2270–2285 (2005)
7. Toh, K.A., Kim, J., Lee, S.: Biometric Scores Fusion Based on Total Error Rate Minimization. *Pattern Recognition* 41, 1066–1082 (2008)
8. Chatzis, V., Bors, A.G., Pitas, I.: Multimodal decision-level fusion for person authentication. *IEEE Transaction on Systems, Man and Cybernetics* 29, 674–680 (1999)
9. Prabhakar, S., Jain, A.K.: Decision-level Fusion in Fingerprint Verification. *Pattern Recognition* 35, 861–874 (2002)
10. Yao, Y.F., Jing, X.Y., Wong, H.S.: Face and Palmprint Feature Level Fusion for Single Sample Biometrics Recognition. *Neurocomputing Letters* 70, 1582–1586 (2007)

11. Yang, J., Yang, J.Y., Frangi, A.F.: Combined Fisherfaces framework. *Image and Vision Computing* 21, 1037–1044 (2003)
12. Yang, J., Yang, J.Y., Zhang, D., Lu, J.F.: Feature Fusion: Parallel Strategy vs. serial strategy. *Pattern Recognition* 36, 1369–1381 (2003)
13. Turk, M., Pentland, A.: Face Recognition Using Eigenfaces. In: *Proc. IEEE Computer Vision and Pattern Recognition*, pp. 586–591 (1991)
14. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 19, 711–720 (1997)
15. Daugman, J.: How Iris Recognition Work. *IEEE Transaction on Circuits and Systems for Video Technology* 14, 21–30 (2004)
16. Ma, L., Tan, T., Wang, Y., Zhang, D.: Personal Identification Based on Iris Texture Analysis. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 25, 1519–1533 (2003)
17. CASIA Iris Image Database,
<http://www.cbsr.ia.ac.cn/english/Databases.asp>
18. ORL Face Image Database, <http://www.cam-orl.co.uk>
19. Yale Face Image Database,
<http://cvc.yale.edu/projects/yalefaces/yalefaces.html>
20. ISO/IEC 19795-1:2006 Information technology - Biometric Performance Testing and Reporting - Part 1: Principles and Framework

Watermark Image Restoration Method Based on Block Hopfield Network

Xiaohong Ma¹, Xin Li¹, and Hualou Liang²

¹School of Electronic and Information Engineering, Dalian University of Technology,
Dalian, 116023, People's Republic of China
maxh@dlut.edu.cn

²School of Biomedical Engineering, Drexel University, 3141 Chestnut Street,
Philadelphia, PA 19104, USA
hualou.liang@drexel.edu

Abstract. In this paper, Hopfield network is introduced for the restoration of extracted watermark image which may be blurred due to the signal transfer or various signal processing operations. A novel codebook method is designed to reduce the storage space of the network and to increase the security. First, each watermark image is divided into adjacent and non-overlapped sub-block images and mapped into a codebook. Second, this codebook is encrypted by a chaotic sequence. During the process of watermark restoration, the codebook can be obtained via a secret key which is then used to construct block weight matrix of the neural network for the restoration of the blurred watermark images. Simulation results demonstrate the excellent performance of the proposed method.

Keywords: Watermark image restoration, Neural network, Block Hopfield network, Codebook.

1 Introduction

Watermarking techniques involve the embedment of information into a digital signal [1-3], and the transmission of this information to the receiver with minimum distortion. The techniques have been increasingly used to prevent or deter the illegal copying, forgery and distribution of digital audio, images, video, and even software. Watermark image, for example, is a common form of watermarks [4,5], where the image used for authentication can be a specific symbol which usually represents the important copy-right information. However, if the extracted watermark image is blurred and unrecognized, the watermarking methods take no effect at all.

While many watermarking schemes have been extensively studied [6-8], none has been focused on watermark restoration method. In this paper, the Hopfield network [9] is introduced to watermark restoration technique. To reduce the storage space of the network consumed, the codebook method is designed. First, each of the watermark images is divided into adjacent and non-overlapped sub-block images and mapped into a codebook. Second, this codebook is encrypted by a chaotic sequence. During the process of watermark restoration, with the help of secret key, the codebook can be

obtained. This codebook is then used to construct block weight matrix of the neural network which helps restore the blurred watermark images. Experimental results showed that the proposed method with Hopfield network renders the watermark images with good visual quality and saves considerable storage spaces.

2 Algorithm

The proposed restoration system, shown in Fig.1, is built upon a watermark system, consisting of the following five components: A. preprocessing watermark images, B. mapping codebook, C. encrypting codebook, D. constructing block weight matrix, and E. restoring blurred images.

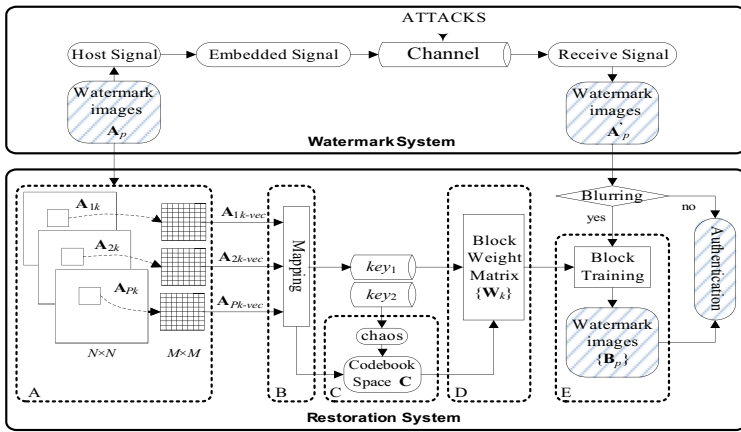


Fig. 1. Watermark system and restoration system

2.1 Preprocessing Watermark Images

For the multiple watermark technique, there are usually two or more watermarks. Here, P binary watermark images, which are denoted as $A_p, p=1,2, \dots, P$ with size $N \times N$, are taken as an example. As shown in Fig. 1, each watermark image is divided into $M \times M$ size of adjacent and non-overlapped sub-block images $A_{pk}, p=1,2, \dots, P, k=1,2, \dots, N^2/M^2$, where k is the number of sub-blocks.

2.2 Mapping Codebook

For a Hopfield network, the weight matrix is typically stored in the trained network, which always takes considerable storage space. To overcome this shortcoming, a method is proposed in this paper to store the codebook instead of the trained weight matrix. Specifically, each $M \times M$ sub-block A_{pk} is first converted into a column vector $A_{pk-vec} = [a_{pk}(1), a_{pk}(2), \dots, a_{pk}(M^2)]^T$. A codebook $C = \{A_{pk-subset}\}$ is then constructed via a unique subset of A_{pk-vec} , and the label of each A_{pk-vec} in C is retained as secret key,

namely key_1 . For different p and k , \mathbf{A}_{pk-vec} may be the same, yet for the codebook \mathbf{C} , all the column vectors $\mathbf{A}_{pk-subset}$ are rarely the same.

2.3 Encrypting Codebook

The codebook, as a public key, will be used in the watermark extraction procedure. As such, certain security methods should be employed. Here, a chaotic sequence [10] is adopted to encrypt codebook \mathbf{C} . The parameter used to generate the sequence is kept as the secret key, namely key_2 .

2.4 Constructing Block Weight Matrix

In the process of watermark extraction, the extracted watermark images are first tested to determine the degree of blurring. If they are blurred with difficulty in authentication, the proposed restoration system will be taken effect. First, the encrypted codebook \mathbf{C} is decrypted using the same chaotic sequence with the help of key_2 . Second, with the label information key_1 and decrypted cookbook, block weight matrix $\mathbf{W}_k, k=1,2, \dots, N^2/M^2$ of each sub-networks can then be reconstructed for the restoration of blurred images as:

$$\mathbf{W}_k = \sum_{p=1}^P (\mathbf{A}_{pk-vec} \cdot \mathbf{A}_{pk-vec}^T - \mathbf{I}) \tag{1}$$

where \mathbf{I} is an identity matrix.

2.5 Restoring Blurred Images

The extracted watermark images are assumed to be $\mathbf{A}'_p = \mathbf{A}_p + \Delta_p, p=1,2, \dots, P$, where Δ_p reflects the difference between original image \mathbf{A}_p and extracted image \mathbf{A}'_p . Following the same codebook processing procedure as module A and B in Fig. 1, the image \mathbf{A}'_p is divided into $M \times M$ size of adjacent and non-overlapped sub-block images \mathbf{A}'_{pk} and converted to a column vector $\mathbf{A}'_{pk-vec} = [a'_{pk}(1), a'_{pk}(2), \dots, a'_{pk}(M^2)]^T$. Assuming $\mathbf{B}_{pk}(0) = \mathbf{A}'_{pk-vec}$, the following iterative equation is adopted to restore the sub-block image:

$$\mathbf{B}_{pk}(t+1) = \text{sgn}(\mathbf{W}_k \cdot \mathbf{B}_{pk}(t)), t = 0,1,2 \dots \tag{2}$$

until $\mathbf{B}_{pk}(t+1) = \mathbf{B}_{pk}(t)$. Here, $\text{sgn}(\cdot)$ is the signum function. Up to this point, $\mathbf{B}_{pk}(t)$ is the restored vector of \mathbf{A}'_{pk-vec} . All restored sub-block image vectors $\mathbf{B}_{pk}(t)$ are combined together to generate the restored watermark image \mathbf{B}_p , which is the restoration of the image \mathbf{A}'_p .

3 Experimental Results

Extensive computer simulations are carried out to demonstrate the validity of the proposed method. The parameters in the simulation are set as follows: $N=32, M=8$, and $P=3$. The three original watermark images are shown in Fig. 2.

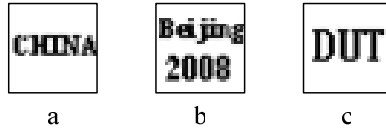


Fig. 2. Three watermark images. a. watermark image 1. b. watermark image 2. c. watermark image 3.

In watermarking technique, severe attacks usually cause serious visual distortion to watermarking images [4]. The blurred images, shown in Figs. 3a, c and e, respectively, are three examples of them. Figs. 3b, d and f are the restored images of good quality with the proposed block Hopfield network.

To quantify the difference between original watermark and extracted blurred/restored watermark image, a Normalised Hamming Distance (*NHD*) is adopted:

$$NHD = \frac{1}{N \times N} \sum_{n1=1}^N \sum_{n2=1}^N \mathbf{A}(n1, n2) \oplus \mathbf{A}'(n1, n2) \in [0, 1]. \tag{3}$$

where \mathbf{A} is the original clear watermark image, \mathbf{A}' is the extracted blurred/restored watermark image, and \oplus represents the XOR operation. The larger the *NHD* is, the more different the two images are.

The *NHDs* for the watermark images in Figs. 3a, c and e are 0.1621, 0.1738 and 0.1963, respectively, whereas *NHDs* are all zeros for images in Figs. 3b, d and f, indicating that the restored images have the same quality as the original watermark images.

Comparison of the *NHDs* between the restored images and that of the blurred images for watermark image 3, as shown in Fig. 4, reveals three distinct observations: (1) if the *NHD* of extracted blurred image is below 0.2441, the watermark image can be restored without any distortion; (2) if the *NHD* of extracted blurred image is above 0.2559 and below 0.3789, it can make the blurred image clear with some negligible distortion, which nevertheless doesn't affect the image content identification; and (3) if the *NHD* of extracted blurred image is above 0.4023, the restored image can still be extracted with some visible noise effects.



Fig. 3. Extracted blurred images and restored images. a. blurred watermark image 1, *NHD* is 0.1621. b. restored watermark image 1, *NHD* is 0. c. blurred watermark image 2, *NHD* is 0.1738. d. restored watermark image 2, *NHD* is 0. e. blurred watermark image 3, *NHD* is 0.1963. f. restored watermark image 3, *NHD* is 0.

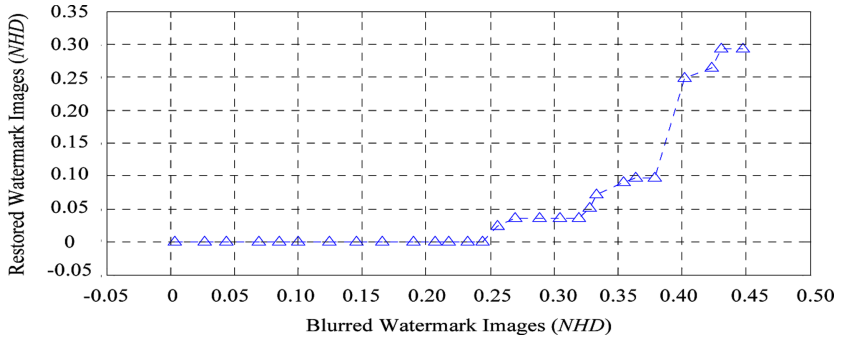


Fig. 4. Restoration results of blurred watermark images 3 with different *NHD*s

To visually appreciate the difference of the above-mentioned three observations, Fig. 5 shows the examples of the blurred and the restored images for each observation in which Fig. 5a, c, and e are the blurred images, and Fig. 5b, d, and f are the corresponding restored image results. We can see from Fig. 5 the clear difference of the quality of the restored images. These results therefore further confirm the data in Fig. 4.



Fig. 5. Three restored results for blurred image 3. a. blurred image, *NHD* is 0.1240. b. restored image, *NHD* is 0. c. blurred image, *NHD* is 0.3047. d. restored image, *NHD* is 0.0352. e. blurred image, *NHD* is 0.4023. f. restored image, *NHD* is 0.2490.

Table 1. Comparison of storage space between our proposed method and the weight matrix method

Block (Yes/No)	No	Yes
Store Style	Weight Matrix	Codebook
Storage Size	N^4	$\leq P N^2$

In addition, the proposed method can greatly reduce the storage space. Comparison of storage space by using our method and by the weight matrix method is shown in Table 1. It can be seen from Table 1 that the storage space is significantly saved with the proposed method.

4 Conclusion

In this paper, the block Hopfield network is introduced for the watermark images restoration. First, the watermark images are mapped into a codebook. Second, it is used to

construct block weight matrix of the neural network which is used to restore the blurred watermark images. In addition, the codebook mapping method can save significant storage space. Experimental results showed that the proposed method can restore blurred watermark images with good quality and save considerable storage space.

Acknowledgements. This work is supported by the National Natural Science Foundation of China under Grant No. 60575011 and Liaoning Province Natural Science Foundation of China under Grant No. 20052181.

References

1. Yen, E., Tsai, K.S.: HDWT-based Grayscale Watermark for Copyright Protection. *Expert Systems with Applications* 35, 301–306 (2008)
2. Tsai, H.H., Sun, D.W.: Color Image Watermark Extraction Based on Support Vector Machines. *Information Sciences* 177, 550–569 (2007)
3. Lin, S.D., Kuo, Y.C., Yao, M.H.: An Image Watermarking Scheme with Tamper Detection and Recovery. *International Journal of Innovative Computing, Information and Control* 3, 1379–1387 (2007)
4. Lin, W.H., Horng, S.J., Kao, T.W., Fan, P., Lee, C.L., Pan, Y.: An Efficient Watermarking Method Based on Significant Difference of Wavelet Coefficient Quantization. *IEEE Transactions on Multimedia* 10, 746–757 (2008)
5. Charalampidis, D.: Improved Robust VQ-based Watermarking. *Electronics Letters* 41, 1272–1273 (2005)
6. Cox, I.J., Miller, M.L., Bloom, J.A., Fridrich, J., Kalker, T.: *Digital Watermarking and Steganography*, 2nd edn. Morgan Kaufmann, San Francisco (2007)
7. Wang, S.S., Tsai, S.L.: Automatic Image Authentication and Recovery Using Fractal Code Embedding and Image Inpainting. *Pattern Recognition* 41, 701–712 (2008)
8. Ma, X.H., Zhang, B., Ding, X.Y.: Self-synchronization Blind Audio Watermarking Based on Feature Extraction and Subsampling. In: Liu, D., Fei, S., Hou, Z., Zhang, H., Sun, C. (eds.) *ISNN 2007*. LNCS, vol. 4492, pp. 40–46. Springer, Heidelberg (2007)
9. Hopfield, J.J.: Neural Networks and Physical Systems with Emergent Collective Computational Abilities. In: *Proceedings of the National Academy of Sciences*, vol. 79, pp. 2554–2558. National Academy Press, USA (1982)
10. Xiang, F., Qiu, S.S.: Analysis on Stability of Binary Chaotic Pseudorandom Sequence. *IEEE Communications Letters* 12, 337–339 (2008)

An English Letter Recognition Algorithm Based Artificial Immune

Chunlin Liang, Lingxi Peng*, Yindie Hong, and Jing Wang

Software School, Guangdong Ocean Univ., Zhanjiang 524025, China
manplx@163.com

Abstract. In the study of letter recognition, the recognition accuracy is impacted by fonts and styles, which is the main bottleneck that the technology is applied. In order to enhance the accuracy, a letter recognition algorithm based artificial immune, referred to as LEBAI, is presented. Inspired by nature immune system, antibody cell (B-cell) population is evolved until the B-cell population is convergent through the learning of each training antigen and the memory cells pool is updated by the optimal B-cell. Finally, recognition is accomplished by memory cells. It is tested by the well-known letter recognition data set of UCI (University of California at Irvine). Compared with HSAC (Letter Recognition Using Holland-Style Adaptive Classifiers), LEBAI showed that recognition accuracy is increased from 82.7% to 95.58%. LEBAI achieves the same recognition accuracy for the letters of different fonts and styles, or stretched and distorted randomly.

Keywords: Letter recognition, Pattern recognition, Artificial immune system, Machine learning.

1 Introduction

Artificial immune is an intelligent information processing mechanisms simulating by biological immune system. The unique information processing mechanisms of immune is self-adaptive, self-learning, self-organized, parallel processing, and coordinately distributed, which provide a powerful paradigm to solve difficult problems. It has been widely applied to various fields. Many well-known results have been achieved in research of combinatorial optimization, machine learning, and etc [1-6].

In the study of letter recognition, the recognition accuracy is impacted by fonts and styles, which is the main bottleneck that the technology is applied. Frey & Slate [7] presented letter recognition using Holland-style adaptive classifiers as HSAC on the basis of Holland's study. HSAC is a letter recognition algorithm, which can recognize letter images that are generated by randomly distorting pixel images of the 26 uppercase letters from 20 different commercial fonts and 6 different letter styles. HSAC

* Corresponding author.

shows lower recognition accuracy. Fogarty [8] presented first nearest neighbor classification on Frey and Slate's letter recognition problem as FNNC. FNNC shows higher recognition accuracy, although the recognition accuracy is increased at the great cost of increasing sharply recognition rules.

In addition, Zhu Li et al. [9] presented a character recognition adaptive learning algorithm based on SVM and sigmoid function. The recognition accuracy of adaptive data is increased by amending self- adaptive the parameters of the sigmoid function, such that the sigmoid function to better fit the class posterior probability distribution of adaptive data output distance. Li Xu et al. [10] presented a container's character recognition algorithm based on neural networks. This algorithm combines several recognizers together based on the peculiarity of each one through the series and parallel hybrid plan, so that increases the recognition accuracy. Wu Ling-chao et al. [11] presented a character recognition based on independent component analysis. The algorithm combines the distance of Euclidean and Mahalanobis based on the principles of independent component analysis to implement character recognition. Li Zuo et al. presented a character recognition approach based on feature line necessary-sufficient condition detection [12]. The algorithm recognizes character through extracting the feature lines from bitmap of character and detecting the necessary-sufficient condition with templates.

Overall, the recognition accuracy of these algorithms is higher, which is limited to the standards character for a kind of font and style. As regards the characters of different fonts and styles, or stretched and distorted randomly, these algorithms are very inefficiently.

In order to enhance the recognition accuracy for the letter with different fonts and styles, a letter recognition algorithm based artificial immune, referred to as LEBAI, is presented. It is tested by the well-known letter recognition data set of UCI (University of California at Irvine) [13] and compared with HSAC etc. LEBAI shows that the recognition accuracy is increased. Furthermore, learning items obvious is reduced and the rate of learning convergence is improved. It also achieved the same accuracy for letters with different fonts and styles, even randomly stretched and distorted.

2 Proposed Algorithm

The principle of LEBAI is based on immune response of organism to antigen. First, the immune system of organism responded to the antigen, and extracted the characteristics of antigen by antibody cell (B-cell) when the organism is attacked. Afterwards, the B-cell is cloned and mutated. The B-cell competed with other in population. The optimal B-cell what the affinity is higher with the antigen is reserve and became the memory cell with the longer life cycle. Finally, the immune system can respond rapidly to the same antigen by the memory cell.

In the algorithm of LEBAI, the training data is equivalent of intrusion antigen. The clone amount and mutation probability of B-cell is adjusted dynamically according to the affinity between B-cell and antigen. On one hand, the lower affinity with the antigen, the smaller B-cell is stimulated; meantime, the fewer amount of cloning, the greater the probability of mutation. On the other hand, the B-cell is greater stimulated,

the more the amount of cloning, and the smaller the probability of mutation. After B-cell is cloned and mutated, the B-cells that can recognize the foreign antigen had a longer life cycle, which would become the memory cells.

LEBAI first learn each antigen of training data one by one and evolved a stable memory cells. Finally, the recognition of antigen is accomplished by memory cells.

Before the introduction the algorithm of LEBAI, the terminologies, symbols, as well as the formulas are defined first.

Tuple $\langle f, c, t \rangle$ is defined as artificial immune cell, referred to as *AIC*. *AIC* is composed of antigens set as *AG*, artificial recognition antibodies set as *AB*, and memory cells set as *MC*, such that $AG \cup ARB \cup MC = AIC$. f is defined as the feature vector of in $\langle f, c, t \rangle$, such that f_i is the same as that the i value of the feature vector, where f_i is real number, $i = \{1, 2, \dots, L\}$, and L is a natural number of dimension of the feature vector. C is defined as a set to express all classes of antibody, and c is a positive integer to express the one of all classes, such that $c \in C = \{1, 2, \dots, nc\}$, where nc is the size of C . t is defined as a life cycle of antibody. In addition, ag , ab , and mc as were defined as an antigen, a *ARB* cell, and a memory cell, respectively.

The affinity of antibody with antigen is based on the similarity of each structure, which is associated with their distance. The fewer the distance, the higher the affinity they have. Let $D(ag, ab)$ represent the distance of antibody with antigen, which is defined as Eq.(1).

$$D(ag, ab) = \sqrt{\sum_{i=1}^L |ag \cdot f_i - ab \cdot f_i|} \tag{1}$$

Let $affinity(ag, ab)$ represent the affinity of antibody with antigen (see Eq.(2)), such that $affinity(ag, ab) \in (0, 1]$.

$$affinity(ag, ab) = \frac{1}{1 + D(ag, ab)} \tag{2}$$

2.1 Initialization

First of all, the characteristics value of the feature vector of training antigens are standardized, which lead to the matrix. The matrix is defined as Eq. (3), where n is the number of training antigens.

$$AG = \begin{pmatrix} ag_1 \cdot f_1 & ag_1 \cdot f_2 & \dots & ag_1 \cdot f_L \\ ag_2 \cdot f_1 & ag_2 \cdot f_2 & \dots & ag_2 \cdot f_L \\ \vdots & \vdots & & \vdots \\ ag_n \cdot f_1 & ag_n \cdot f_2 & \dots & ag_n \cdot f_L \end{pmatrix} \tag{3}$$

Then, the algorithm selected randomly m antigens to take shape the initial set of artificial antibodies and memory cells.

2.2 The Clone and Mutation of Antibody Cell

After the initialization accomplished, LEBAI then learn from each of the training antigens. First, the memory cell as mc_{match} is found from memory cells based on Eq. (4), where the class of mc_{match} is the same as the class of the learning antigen, and the $affinity(ag, ab)$ is highest. If the mc_{match} is not found, $mc_{match}=ag$, and ag will be added to the memory cells set.

$$mc_{match} = \begin{cases} ag & \text{iff } MC_{ag.c} = \varphi \\ \max(affinity(ag, mc)) & MC_{ag.c} \neq \varphi \end{cases} \quad (4)$$

Afterwards, the mc_{match} is cloned. The amount of cloning, referred to as $cCount$, is based on the affinity of mc_{match} with the learning antigen. The higher the affinity, the mc_{match} is stimulated greater, the more the amount. The $cCount$ is decided by the two parameters, the one is the stimulated value of cloning as $cStim$, the other one is the constant of cloning as $cConst$. The $cStim$ is defined by Eq. (5). The $cConst$ is a input constant, where the $cConst$ is used to ensure that the new B-cells were enough to be added to B-cells. The $cConst$ is defined by Eq. (6).

$$cStim = \frac{1 + affinity(ag, mc_{match})}{2} \quad (5)$$

$$cCount = cStim * cConst \quad (6)$$

Finally, the feature vector of the new B-cells is mutated. The probability of mutation, referred to as $mRate$, is based on the affinity of mc_{match} with the learning antigen. The higher the affinity, the mc_{match} is stimulated fewer, and the lower the probability. The $mRate$ is decided by the two parameters, the one is the stimulated value of mutation as $mStim$, and the other one is the constant of mutating as $mConst$. The $mStim$ is defined by Eq. (7). The $mConst$ is a positive constant, where the $mConst$ is used to adjust the tempo that the eigenvector is mutated. The $mRate$ is defined by Eq. (8).

$$mStim = \frac{1}{1 + affinity(ag, mc_{match})} \quad (7)$$

$$mRate = mStim * mConst \quad (8)$$

The mutated B-cells will be added to B-cells after the mc_{match} has been cloned and mutated.

2.3 The Controlling of Antibody Cell Scale

After the antibody cells have been cloned and mutated, the scale of antibody cells will be expanded rapidly. When the amount of antibody cells reached a certain threshold, in order to control the scale of antibody cells and the convergence of the algorithm, some of the antibody cells would be eliminated through competition.

The life cycle of the antibody cell t_{ab} and right weight w_{ab} are defined by Eq. (9) and Eq. (10), respectively. The larger w_{ab} , the longer life cycle t_{ab} has. p is the amount of the antibody cells which has the same class as the learning antigen. t_0 is the life cycle of the antibody and the initial t_0 is 0, and The $tConst$ is integer constant, where the $tConst$ is used to expand the life cycle scale of the antibodies with same class.

$$w_{ab} = \frac{\sum_{j=1, j \neq i}^{j=p} D(ab_i, ab_j)}{p-1} \quad \text{iff } ab_j \in AB_i \text{ and } ab_i.c = ab_j.c \quad (9)$$

$$t_{ab} = t_0 + tConst * w_{ab} \quad (10)$$

Afterwards, some of the antibody cells would be eliminated based on the life cycle of the antibody cell. The process is as follows:

- (1) Update the life cycle of the antibody cell based on Eq. (10), where the class of the antibody cell has the same as the learning antigen.
- (2) Compute the average of the max life cycle, referred to as avg_{max} , based on Eq. (11), where the max life cycle has the highest value in the antibody cells of the same class.
- (3) Compute the average of the min life cycle, referred to as avg_{min} , based on Eq. (12), where the min life cycle has the lowest value in the antibody cells of the same class.
- (4) If $t_{ab} > avg_{max}$, the antibody cell is eliminated by aging; If $t_{ab} < avg_{min}$, the one will be eliminated by life weak. The antibody cells between avg_{min} and avg_{max} will be reserved to become the candidate memory cell.

$$avg_{max} = \frac{\sum_{c=1}^{nc} ab_c \cdot t_{max}}{nc} \quad \text{iff } c \in C \quad (11)$$

$$avg_{min} = \frac{\sum_{c=1}^{nc} ab_c \cdot t_{min}}{nc} \quad \text{iff } c \in C \quad (12)$$

2.4 Updating of Memory Cells

The end of the algorithm is to select the memory cells from the antibody cells. If an antibody cell can recognize the learning antigen, it would become the candidate memory cell and be added to the memory cells set to replace the one of the memory cells, on condition that the candidate memory is better than the one. The process leads to a convergence memory cells set and stable immune system, which will have a rapid second response to intrusion antigen. The process is as follows:

$$mc_{cand} = \max(\text{affinity}(ag, ab)) \quad \text{iff } ag.c = ab.c \quad (13)$$

First, the algorithm selects the candidate memory cell based on Eq. (13) from the antibody cells.

$$mc_{replaced} = \max(\text{affinity}(ag, mc)) \text{ iff } ag.c = mc.c \quad (14)$$

Afterwards, the algorithm selects the replaced memory cell based on Eq. (14) from the memory cells.

Finally, if the affinity of the learning antigen with the mc_{cand} is higher than with $mc_{replaced}$, $mc_{replaced}$ will be replaced by mc_{cand} .

2.5 Recognition

After the end of the learning, the class of testing antigen will be decided by the memory cell that has the highest affinity with the testing antigen. The process is as follows:

1. Compute the affinity of each one of memory cells with the testing antigen based on Eq. (2).
2. Select the memory cell which has the highest affinity with the testing antigen. The class of the testing antigen has the same class with the memory cell.

3 Experiments

3.1 Dataset

The experiment used the well-known letter recognition dataset of UCI (University of California at Irvine) to test the recognition performance of LEBAI [13]. The data set comprised 20,000 learning items. The data of each learning item came from the different image of the letters.

3.2 Experiment Parameters

The experiment selects randomly the 4000 learning item as the test set from the data set, the other 16000 learning item as a training set. Table 1. show the parameters.

Table 1. Parameters

Parameter	Value
cConst	1000
mConst	0.5
tConst	3000
nc	26

3.3 Experiment Results

Different sizes of memory cells were adopted to test the recognition accuracy, which is shown in Fig. 1. LEBAI shows that the greater the size of memory cells, the higher the recognition accuracy. The ultimate recognition accuracy achieved maximum 95.58 %.

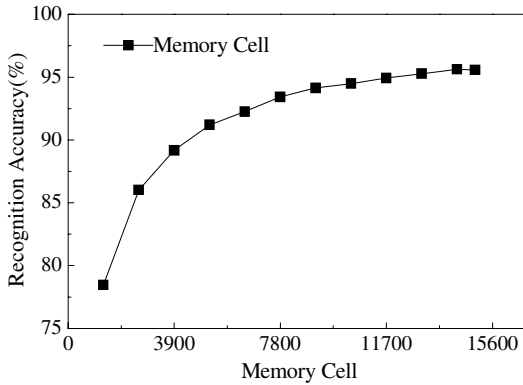


Fig. 1. The size of memory cells to recognition accuracies

Different sizes of training antigens were adopted to test the recognition accuracy, which is shown in Fig.2 where sizes of memory cell are 14850, 10000, and 5200, respectively. The experiments show that with the increase of training antigen, the recognition accuracy is increased. However, when the size of memory cell is over 10000, the experiment results is stable to 95.87% because the convergence of the memory cell.

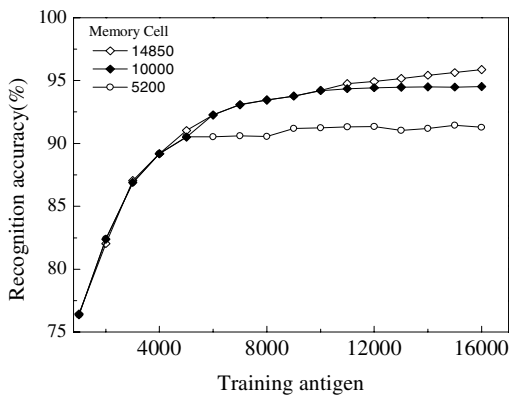


Fig. 2. Training antigens to recognition accuracies

Table 2. The comparison of recognition accuracy

Algorithm	Accuracy (%)
HSAC [7]	82.70
Genetic programming [14]	92.00
Neural networks [15]	94.13
FNNC [8]	95.67
LEBAI	95.87

In order to prove that LEBAI improves the recognition accuracy, the comparison of recognition accuracy is showed in Table.2, compared with FNNC, HSAC and etc recognition algorithms. LEBAI shows that the recognition accuracy achieves the highest recognition accuracy.

4 Conclusion

The reasons that the existed letter recognition algorithms have low recognition accuracy are analyzed. Afterwards, a novel letter recognition algorithm based artificial immune is presented. Compared with some well-known letter recognition algorithms, LEBAI shows higher recognition accuracy.

References

1. Albert, R., Jeong, H., Barabasi, A.: Attack and Error Tolerance of Complex Networks. *Nature* 406, 378–382 (2002)
2. Li, T.: Dynamic Detection for Computer Virus based on Immune System, *Science In China. Series F: Information Science* 51(10), 1475–1486 (2008)
3. Omkar, S., Khandelwal, R., Yathindra, S., et al.: Artificial Immune System For Multi-Objective Design Optimization of Composite Structures. *Engineering Applications of Artificial Intelligence* 21(8), 1416–1429 (2008)
4. Vijayalakshmi, K., Radhakrishnan, S.: Artificial Immune based Hybrid GA for QoS based Multicast Routing in Large Scale Networks (AISMR). *Computer Communications* 31(17), 3984–3994 (2008)
5. Wang, L., Singh, C.: Population-based Intelligent Search in Reliability Evaluation of Generation Systems with Wind Power Penetration. *IEEE Transactions on Power Systems* 23(3), 1336–1345 (2008)
6. Ye, F., Xu, S., Xiong, Y.: Two-step Image Registration by Artificial Immune System and Chamfer Matching. *Chinese Optics Letters* 6(9), 651–653 (2008)
7. Frey, P.W., Slate, D.J.: Letter Recognition Using Holland-style Adaptive Classifiers. *Machine Learning* 6(2), 161–182 (1991)
8. Fogarty: First Nearest Neighbor Classification on Frey and Slate's Letter Recognition Problem. *Machine Learning* 9(4), 387–388 (1992)
9. Zhu, L., Sun, G.: Character Recognition Adaptive Learning Algorithm based on SVM and Sigmoid Function. *Application of Electronic Technique* 32(4), 16–17 (2006)
10. Li, X., Yang, J.: Container's Character Recognition Algorithm based on Neural Networks. *Computer and Communications* 19(z1), 89–91 (2001)

11. Wu, L., Mo, Y.: Character Recognition based on Independent Component Analysis. *Journal of Shanghai University* 9(3), 193–196 (2003)
12. Li, Z., Wang, S., Cai, S.: Character Recognition Approach based on Feature Line necessary-sufficient condition detection. *Journal of Software* 13(1), 85–91 (2002)
13. Frey, P.W., Slate, D.J.: UCI Repository of Machine Learning Databases, Letter Recognition Datasets (1991),
<http://archive.ics.uci.edu/ml/datasets/Letter+Recognition>
14. Ahluwalia, M., Bull, L.: Coevolving Functions in Genetic Programming. *Systems Architecture* 47 (2001)
15. Daqi, G., Chao, X., et al.: Combinative Neural-network-based Classifiers for Optical Handwritten Character and Letter Recognition. *International Joint Conference on Neural Networks*, 3, 2232–2237 (2003)

Interpretation of Ambiguous Zone in Handwritten Chinese Character Images Using Bayesian Network

Zhongsheng Cao, Zhewen Su, and Yuanzhen Wang

College of Computer Science and Technology,
Huazhong University of Science and Technology, Wuhan 430074, China
{caozhongsheng,szwc1ever,wangyz2005}@163.com

Abstract. Interpretation of ambiguous zone is an essential step to recovering dynamic information from handwritten images, which can be seen as to deduce the original motion intention of the writer at the intersection areas. This study presents a novel method to interpret ambiguous zones by constructing a Bayesian belief network. In the initial phase, a graph is built to model the character and several sample points are extracted from each sub-stroke. In the interpreting phase, each pair of sub-strokes is characterized in terms of the comparison of orientation, width, and curvature. Finally, a Bayesian belief network is established to determine the continuous pairs. A series of experiments are conducted on test samples collected from a standard handwritten Chinese text database, and the results show that the proposed method can interpret ambiguous zones effectively.

Keywords: Bayesian network, Handwritten Chinese character, Ambiguous zone, Stroke extraction, Handwriting. recognition.

1 Introduction

Ambiguous zones are the parts of a character image with the intersections of stroke, where the *odometric* information has been lost or ambiguous and any explicit clues about the original writing trajectory cannot be acquired [1], as shown in Fig. 1(a). Generally, artifacts or pattern distortions are always generated in the ambiguous zones after thinning [2] (see Fig. 1(b)), which lead to erroneous interpretations of strokes in the procedure of extracting strokes [3] or recovering dynamic information [4,5]. Therefore, an appropriate way to deal with ambiguous zones can facilitate the procedures above, which is always divided into two steps: ambiguous zone detection and interpretation.

Several algorithms of ambiguous zone detection have been proposed until now, e.g., window-scanning[1], regularity and singularity analysis[2], constructing CSSG [3], et al. Besides, we have proposed a novel method for detecting ambiguous zones using feature points of the skeleton and the contour information around them in our previous work [6]. In this paper, we focus on the problem of

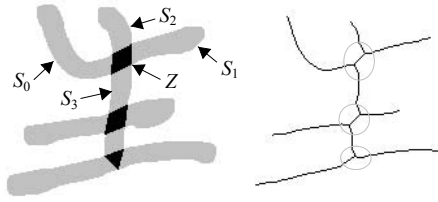


Fig. 1. (a) Ambiguous zones (the black areas) and sub-strokes (the gray areas) in the handwritten Chinese character; (b) distortions in ambiguous zones after thinning

ambiguous zone interpretation. After all the ambiguous zones detected, the character is divided into two parts: ambiguous zones and sub-strokes (see Fig. 1(a)), and interpretation of ambiguous zone is to determine whether or not a pair of sub-strokes jointed at the same ambiguous zone belongs to the same stroke, which plays an important role in the process of character skeletonization [2], stroke extraction [3,6] or dynamic information recovery [1,4,5]. For example, there are four sub-strokes S_0 , S_1 , S_2 and S_3 jointed at the ambiguous zone Z in Fig. 1(a), and (S_0, S_1) and (S_2, S_3) should be continuous pairs, since they are generated by coherent movements.

Generally, previous studies in this area can be divided into two main categories: local analysis [1,3] and global search [7]. In the former, the determination of interpretation is made based on current local configuration. Plamondon and Privitera [1] exploited three parameters involved in the human visual processing of handwritten graphic forms for a correct and plausible interpretation of ambiguous zones. Lee and Wu [3] fitted sample points of each pair of sub-strokes with a Bezier curve, and the interpretation decision was made by estimating the error distance between the fitted curve and the original sub-stroke segments. On the other hand, the global search is conducted in global sense with an evolution function maximized or minimized. Jager [7] exploited a graph model to represent each thinned character and searched a path in the graph based on the *minimum energy cost* criterion. The main drawback of local analysis is the difficulty of designing general heuristic rules which are applicable to various writing styles, and the methods of global search suffer from huge computational cost as well. Although Qiao et al. [4,5] tried to combine the two paradigms together, the method is still limited to the single-stroke images. Besides, some methods also made use of prerecorded dynamic exemplars for the ambiguity interpretation, like [8]. However, the assumed availability of dynamic exemplars may not always hold in practice.

The interpretation of ambiguous zone can be treated as to retrieve a subset of the set of candidate pairs generated by combining each sub-stroke jointed at the ambiguous zone with another sub-stroke. One intuitive solution is to assume that the direction of a segment is maintained when passing through an ambiguous zone, as in [2]. However, it is extremely difficult to obtain the unique solution fully conformed to the real writing trajectory, especially when a certain number

of ambiguous zones gather in a small area and some of the sub-strokes delimited by those ambiguous zones are not long enough to give a correct interpretation. In this paper, we treat the ambiguity interpretation as a classification problem, and utilize a Bayesian belief network to resolve it. The method of statistical analysis is a proper way to estimate continuous pairs, since it can be performed in a systematic and trainable manner to adapt to various writing styles with limited computational complexity. The proposed approach is presented in two phases. In the initial phase, an undirected graph is built to model each character, and sample points of each sub-stroke are extracted by tracing along both boundary courses of the sub-stroke contour. In the interpreting phase, several features are extracted in terms of orientation difference, width comparison, and curvature variation, and the most likely Bayesian network structure is selected based on a scoring criterion function. The Bayesian network is trained by the samples chosen from a standard handwritten Chinese text database randomly, and the outputs give us an encouraging result.

The rest of the paper is organized as follows. Section 2 briefly describes the construction of a graph model and extraction of sample points from sub-strokes. The procedure of interpretation of ambiguous zones is introduced in Section 3. The experimental results and discussions are given in Section 4. Finally, the conclusions of the paper are presented in the last section.

2 Character Modeling

We assume the images considered in this paper have been denoised and binarized. Using an algorithm of ambiguous zone detection, a character is divided into two parts: ambiguous zone and sub-stroke. Here, sub-strokes are defined as the stroke segment separated by a series of ambiguous zones. In this section, we will explain how to build a graph from the set of ambiguous zones and sub-strokes, and extract sample points from each sub-stroke for ambiguity interpretation.

2.1 Graph Representation

To facilitate the interpreting process below, we use an undirected graph $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ to model each character, where \mathbf{V} and \mathbf{E} are the sets of nodes and edges respectively. Each node represents one of the sub-strokes or ambiguous zones. If an ambiguous zone and a sub-stroke, corresponding to the nodes v_i and v_j respectively, are connected to each other, there is an edge between v_i and v_j , denoted as $(v_i, v_j) \in \mathbf{E}$. When the sub-stroke is connected to the ambiguous zone at its both ends (see, for instance, the sub-stroke g and the ambiguous zone h in Fig. 2(a)), a virtual node has to be appended to avoid the parallel edges between v_i and v_j . The virtual node v_n corresponds to the sub-stroke, and $(v_n, v_j) \in \mathbf{E}$, $(v_n, v_i) \in \mathbf{E}$. An illustration is given in Fig. 2, where k is a virtual node.

In the graph, a node with degree three or more represents an ambiguous zone, while a node with degrees less than three corresponds to a sub-stroke.

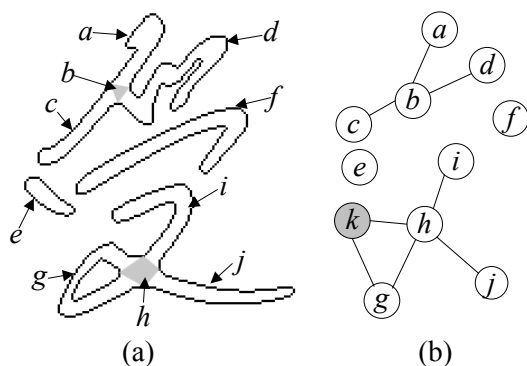


Fig. 2. Graph representation of handwritten Chinese character “Shou”. (a) Sub-strokes (a, c, d, e, f, g, i and j) and ambiguous zones (b and h), (b) corresponding graph.

2.2 Sample Point

We use the method of tracing the double boundary course to obtain sample points of sub-strokes, and the tracing procedure is conducted by two pointers p_0 and p_1 on the both sides of the sub-strokes. Suppose a sub-stroke is connected to an ambiguous zone at the contour point s_0 and s_1 , and s_0 and s_1 are taken as the starting positions of p_0 and p_1 , respectively. Then, move forward the current pointers p_0 and p_1 by the distance between them to obtain the next candidate pointers p'_0 and p'_1 . Let $D(p_0, p_1)$ be the distance between p_0 and p_1 , and $d_0 = D(p'_0, p'_1)$, $d_1 = D(p'_0, p_1)$, $d_2 = D(p_0, p'_1)$:

- If $d_0 \leq d_1$ and $d_0 \leq d_2$, then $p_0 = p'_0$, $p_1 = p'_1$; else
- If $d_1 \leq d_0$ and $d_1 \leq d_2$, then $p_0 = p'_0$; else
- If $d_2 \leq d_0$ and $d_2 \leq d_1$, then $p_1 = p'_1$.

The tracing procedure is shown in Fig. 3. Each time the pointers are relocated, the center point of the straight line between p_0 and p_1 is recorded as a sample point (the filled dots in Fig. 3), and the sequence of sample points will be used to reconstruct strokes in following sections.

One of the major problems of ambiguity interpretation is that some sub-strokes are too short to make a continuity determination. To overcome this problem, those nodes of \mathbf{G} whose corresponding ambiguous zones are very close to each other have to be merged into a new one. Imagine that z_0 and z_1 are two ambiguous zones, connected by a sub-stroke ss . Let S_w be the set of the sample points of ss , if $|S_w| < 3$, then replace the nodes that correspond to z_0 and z_1 with a new node, and the connection relations among these nodes and the others in \mathbf{G} are preserved. In particular, z_0 and z_1 are connected to each other, then $|S_w| = 0$.

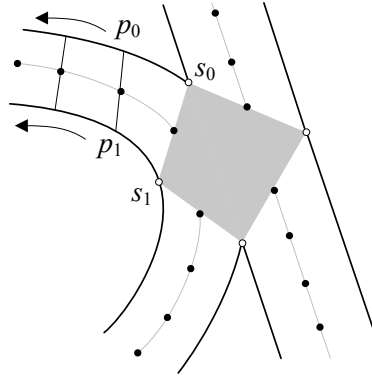


Fig. 3. Tracing along double boundary course of sub-stroke contour

3 Ambiguity Interpretation

In the phase of interpreting, we attempt to deduce the original motion intention of the writer from the detailed information of involved sub-strokes. After ambiguous zone detection, a stroke is considered as the sequence alternated between sub-strokes and ambiguous zones, represented as $(s_0, a_0, s'_0, \dots, s_i, a_i, s'_i, \dots, s_n, a_n, s'_n)$ ($0 \leq i \leq n$), where s_i and s'_i are sub-strokes, a_i is an ambiguous zone, and $s'_i = s_{i+1}$. In global search, the continuity of the pair of sub-strokes (s_i, s'_i) is determined by the evaluation on the whole stroke. However, for a sub-stroke in the sequence s_j , the larger $|i - j|$ is, the less helpful to the estimation of continuity s_j become, or even in some cases the influence is negative. Actually, the detailed information of the joint sub-strokes is convincing enough to make the correct estimation of continuity in most cases. From this sense, one intuitive method is to assign a continuous score for each pair of sub-strokes using a decision function (like [1]). However, it is difficult to design a general decision function. Therefore, in our method, a Bayesian belief network is established to interpret ambiguous zones, in which statistical analysis is exploited to adapt to different handwriting styles.

3.1 Feature Extraction

Basically, interpretation of ambiguous zones is a procedure of the human visual understanding of handmade graphic forms [1]. On the basis of the observation that a continuous pair of sub-strokes is usually generated by a coherent movement, several measurements can be made to estimate the difference between each pair of sub-strokes in this study, including a) stroke width: sub-strokes in a continuous pair should have approximately the same width; b) deviation angle: there is a small deviation angle between a continuous pair of sub-strokes; c) alteration in curvature: human normally writes characters in the smooth way and the curvature cannot be abruptly changed.

Due to the instability of thinned results, we extract features from the sample point sequence of sub-stroke. Suppose v_i is a sub-stroke connected to an ambiguous zone v_a , the sample point sequence of v_i (started from the end connected to v_a) is represented as $Sq_i = (p_{i,0}, p_{i,1}, \dots, p_{i,m})$. The first step to feature extraction is determining a support segment from the sample point sequence, which can be seen as a subsequence of Sq_i starting from the first sample point $p_{i,0}$ (as the segment should be as close as possible to v_a). Thus the length of support segment (the sum of Euclidean distance from point to point) determines the feature resolution. On the one hand, we desire the length of support segment is as short as possible. Because the shorter it is, the closer to our expectation the measurements of feature made on the segment (like tangent direction, curvature, et al.) get. On the other hand, a support segment without large enough length will also cause inaccurate measurements. Based on the discussions above, we give several constraints for the determination of support segment as follows.

Let $V(p_{i,j}, p_{i,k})$ ($0 \leq j, k \leq m, j \neq k$) be the direction angle from $p_{i,j}$ to $p_{i,k}$, and

$$\varphi_i(L) = \frac{1}{L} \sum_{k=1}^L V(p_{i,k-1}, p_{i,k}) (1 \leq L \leq m), \tag{1}$$

$$E(\varepsilon, \eta) = \min(|\varepsilon - \eta|, 2\pi - |\varepsilon - \eta|). \tag{2}$$

The support segment is represented as $(p_{i,0}, p_{i,1}, \dots, p_{i,g_i})$, where g_i ($2 \leq g_i \leq m$) is the maximum integer satisfying the following three conditions:

- a) $\sum_{k=1}^{g_i} d(p_{i,k-1}, p_{i,k}) \leq S_{th}$, where S_{th} is defined as the upper bound of the length of support segment, In the experiment, we found that the value between $2w$ and $3w$ is more suitable for our test samples;
- b) for $k = 1, \dots, g_i$, $E(V(p_{i,k-1}, p_{i,k}), \varphi_i(k)) < \alpha_{th0}$;
- c) for $k = 1, \dots, g_i$, $E(V(p_{i,k-1}, p_{i,k}), V(p_{i,0}, p_{i,1})) < \alpha_{th1}$, where α_{th0} and α_{th1} are the thresholds that restrict the current angel deviation from the mean and the first direction angle, respectively.

After the determination of support segment, we estimate the mean tangent direction of the support segment of v_i as $\alpha_i = \varphi_i(g_i)$ (see Fig. 4). At the same

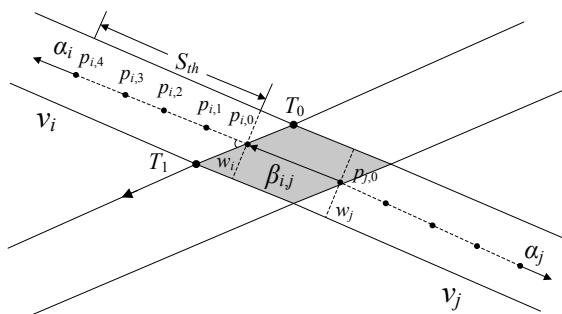


Fig. 4. Estimation of tangent direction and width of sub-stroke

time, the stroke width is taken into consideration as well. Suppose T_0 and T_1 are the vertices of the ambiguous zone v_a which are connected to the contour of v_i (see Fig. 4), the mean stroke width w_i of the support segment can be approximated as

$$w_i = d(T_0, T_1) |\sin(V(T_0, T_1) - \alpha_i)|. \tag{3}$$

For two sub-strokes v_i and v_j connected to the same ambiguous zone v_a , a 5-dimensional feature vector, $X = \{x_0, x_1, x_2, x_3, x_4\}$, is exploited to characterize the attributes as follows:

- 1) x_0 denotes the degree of v_a ;
- 2) x_1 depicts the angle deviation $\gamma_{i,j}$ between v_i and v_j :

$$\gamma_{i,j} = |\alpha_i - \beta_{i,j}| + |\pi - |\alpha_j - \beta_{i,j}||, \tag{4}$$

where $\beta_{i,j} = V(p_{j,0}, p_{i,0})$;

- 3) Let $\gamma_m(i, j)$ be the minimum angle deviation between v_i and other sub-strokes connected to v_a except v_j :

$$\gamma_m(i, j) = \min_{(v_k, v_a) \in E, v_k \neq v_i, v_j} \gamma_{i,k}, \tag{5}$$

and x_2 represents $\min(\gamma_m(i, j), \gamma_m(j, i))$;

- 4) x_3 describes the width difference between v_i and v_j as $w_{i,j} = |w_i - w_j|$;
- 5) x_4 estimates the variation of curvature by

$$\psi_{i,j} = \frac{1}{g_i + g_j - 1} \sum_{k=-g_j}^{g_i-2} \frac{|\zeta_{i,k} - \zeta_{i,k+1}|}{d(p_{i,k}, p_{i,k+1})}, \tag{6}$$

where $\zeta_{i,k}$ is the curvature at the SPWT $p_{i,k}$ of v_i which is calculated by:

$$\zeta_{i,k} = \frac{E(V(p_{i,k-1}, p_{i,k}), V(p_{i,k}, p_{i,k+1}))}{d(c_{i,k-1}, c_{i,k})}, \tag{7}$$

and $c_{i,k}$ is the center point of the straight line between $p_{i,k}$ and $p_{i,k+1}$. $p_{i,-1} = p_{j,0}, p_{i,-2} = p_{j,1}, \dots, p_{i,-g_j-1} = p_{j,g_j}$.

Using the features above, we create a Bayesian belief network for ambiguity interpretation in the following section.

3.2 Bayesian Belief Network

A Bayesian belief network is a graphical model that represents probabilistic dependence relationships among variables of interest. After discretization of the continuous attributes, the feature vector X and the class label attribute C are taken as the system variables. The propositions on node C are described as $c = (c_0, c_1)$ representing the classes “continuous” and “discontinuous”, respectively.

There are many methods of Bayesian structure learning is mentioned in [9], but they all suffer from huge computations. In order to improve the processing speed, we establish several possible structures of the Bayesian network based

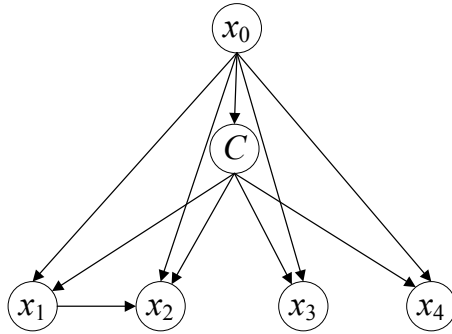


Fig. 5. The Bayesian belief network for ambiguity interpretation

on the understanding of these variables in advance, and select the most likely one with the classical scoring criterion function proposed in [10] maximized (as shown in Fig. 5).

Since the degree of ambiguous zone reflects the local configuration around the area, the other variables are influenced by x_0 . Besides, the feature variables x_1 , x_2 , x_3 and x_4 are influenced by the class label attribute C . Furthermore, there is no correlation between the changes in curvature and width, thus x_4 and x_3 are independent to each other, and obviously the variable x_2 is dependent of x_1 . Given the evidence of all the feature variables, the posterior probability of the proposition $c_i(i = 0, 1)$ on the node C is calculated as

$$\begin{aligned}
 P(c_i|X) &= \frac{P(c_i X)}{\sum_{j=0}^1 P(c_j X)} \\
 &= \frac{P(x_0)P(c_i|x_0)P(x_1, x_2, x_3, x_4|c_j, x_0)}{\sum_{j=0}^1 P(x_0)p(c_j|x_0)P(x_1, x_2, x_3, x_4|c_j, x_0)} \\
 &= \frac{P(c_i|x_0)p(x_2|c_i, x_0, x_1) \prod_{k=1,3,4} P(x_k|c_i, x_0)}{\sum_{j=0}^1 P(c_j|x_0)P(x_2|c_j, x_0, x_1) \prod_{k=1,3,4} P(x_k|c_j, x_0)}. \tag{8}
 \end{aligned}$$

The conditional probability tables for each variable are calculated from a set of manual labeled samples collected from a standard handwritten Chinese character database, and the details will be given in Section 4.

4 Experiments and Discussions

To verify the validity of the proposed method, 1000 handwritten Chinese character images, coming from 50 different writers, were chosen randomly from HIT-MW database [11]. After ambiguous zone detection, we selected four-tenth of

these samples for training the Bayesian belief network, and the other six-tenth for testing. The proposed method was implemented on PC with P4 1.8 GHz processor and 512M memory under the development environment of Visual Studio .NET 2003.

In training phase, each pair of sub-strokes in the set of training samples was classified into two groups manually according to their continuities, and the features were extracted by the procedure discussed previously. There are 2744 pairs of sub-strokes are continuous and 5689 pairs are discontinuous. Then, we calculated the probability assignments associated with all the possible network structures, and the most likely one is chosen by maximizing the scoring criterion function.

In the testing, 17443 pairs of sub-strokes coming from the testing sample set were analyzed, some results are given in Fig. 6, where the double arrows represent continuous pairs, while the single arrows identify sub-strokes that terminated in ambiguous zones. We also compared our method with the methods proposed in [1] and [4]. In [1], local analysis is employed to assign a continuous score for each pair of sub-strokes by an interpretation function, whereas the method in [4] evaluate the local continuity analysis within a probability framework and made the final determination by global smoothness calculation. The comparative results are given in Table 1. It can be concluded the proposed method has the best performance, while the time cost is also acceptable.



Fig. 6. Experiment results

Table 1. Comparison of ambiguity interpretation and time cost

	Our method	Method in [1]	Method in [4]
Accuracy for ambiguity interpretation	95.5%	82.3%	90.1%
Time cost (ms/character)	112.7	88.4	168.4

5 Conclusions

Interpretation of ambiguous zone is a process that determines the right pairs of sub-strokes belonging to the same writing movement, which could effectively improve the performance of the application of skeletonization, stroke extraction and dynamic information recovery in the field of offline handwriting recognition.

In this paper, a novel approach based on Bayesian belief network is presented to analyze the connectivity of each pair of sub-strokes statistically. After ambiguous zone detection, our method is conducted into two phases. In the initial phase, a graph is created from the set of ambiguous zones and sub-strokes, and we extract sample points from each sub-stroke for the continuity analysis. In the interpreting phase, a Bayesian belief network that includes the feature involved in human visual understanding of handwriting is established to classify the candidate pairs of sub-strokes into two groups: continuous and discontinuous. From the experimental results, it can be concluded that the application of the proposed method enables us to interpret ambiguous zones effectively.

References

1. Plamondon, R., Privitera, C.M.: The Segmentation of Cursive Handwriting: An Approach Based on Off-Line Recovery of the Motor-Temporal Information. *IEEE Trans. Image Processing* 8, 80–91 (1999)
2. Zou, J.J., Yan, H.: Skeletonization of Ribbon-Like Shapes Based on Regularity and Singularity Analyses. *IEEE Trans. syst. Man Cybern.* 31, 401–407 (2001)
3. Lee, C., Wu, B.: A Chinese-Character-Stroke-Extraction Algorithm Based on Contour Information. *Pattern Recognition* 31, 651–663 (1998)
4. Qiao, Y., Yasuhara, M.: Recovering Dynamic Information from Static Handwritten Images. In: 9th International Workshop on Frontiers in Handwriting Recognition, pp. 118–123. *IEEE Comput. Soc. Press, Los Alamitos* (2004)
5. Qiao, Y., Nishiara, M., Yasuhara, M.: A Framework Toward Restoration of Writing Order from Single-Stroke Handwriting Image. *IEEE Trans. Pattern Anal. Mach. Intell.* 28, 1724–1737 (2006)
6. Cao, Z.S., Su, Z.W., Wang, Y.Z., Xiong, P.: A Method for Handwritten Chinese Stroke Extraction Based on Ambiguous-Zone Detection. *Journal of Image and Graphics* (accepted) (in Chinese)
7. Jäger, S.: Recovering Writing Traces in Off-Line Handwriting Recognition: Using a Global Optimization Technique. In: 13th International Conference on Pattern Recognition, pp. 150–154. *IEEE Comput. Soc. Press, Los Alamitos* (1996)
8. Nel, E.M., du Preez, J.A., Herbst, B.M.: Estimating the Pen Trajectories of Static Signatures Using Hidden Markov Models. *IEEE Trans. Pattern Anal. Mach. Intell.* 27, 1733–1746 (2005)
9. Cooper, G.F., Herskovits, E.: A Bayesian Method for the Induction of Probabilistic Networks from Data. *Machine Learning* 9, 309–347 (1992)
10. Neapolitan, R.E.: *Learning Bayesian Networks*. Prentice Hall, Upper Saddle River (2004)
11. Su, T., Zhang, T., Guan, D.: Corpus-based HIT-MW Database for Offline Recognition of General-Purpose Chinese Handwritten Text. *Int. J. Doc. Anal. Recognit.* 10, 27–38 (2007)

Weather Recognition Based on Images Captured by Vision System in Vehicle

Xunshi Yan¹, Yupin Luo¹, and Xiaoming Zheng²

¹ Tsinghua National Laboratory for Information Science and Technology (TNList),
Department of Automation, Tsinghua University, Beijing 100084, China

² INF Technologies, Ltd. , Beijing 100086, China

yanxs06@mails.thu.edu.cn, luotsinghua.edu.cn, zheng@tsinghua.edu.cn

Abstract. Weather recognition is widely required in many areas, and it is also a challenging and brand-new subject. This paper proposes an approach to recognize weather based on images captured by in-vehicle vision system. We bring three groups of features, including histogram of gradient amplitude, HSV color histogram, road information, and employ an algorithm based on Real AdaBoost, making use of the category structure to achieve the task of classification. Experiments confirm superior performances on our dataset collected from images captured by vision system.

Keywords: Weather recognition, Vision system, Real AdaBoost, HSV color space, Category structure.

1 Introduction

Computer vision system has achieved great success in many areas, such as surveillance, navigation, driver assistance system. However, the cameras exposed outside are easily influenced by bad weather. For example, pedestrian detection system in vehicle could not work at all when raindrop falls on the camera, even results serious false-detection. Many vision systems also need to reset parameters such as lighting, rain wiper, under different weather conditions. Hence, research of weather recognition in vision system is in urgent demand.

Weather recognition is a brand-new subject and only a few of previous work has addressed this issue. Garg and Naya^[1] in Columbia University focus on detecting and removing rain streaks from videos. The idea comes from moving object detection. It makes a difference between the two adjacent frames, and can give some perfect results under certain scenarios, but it is hard to satisfy the dynamic background or the situation of raindrops adherent on the camera. Kurihata and Takahashi^{[2][3]} in Nagoya University collect large mount of raindrop patches, utilize principal component analysis to make a raindrop template. They search the global or part of images by computing the similarity between the raindrop template and the patches in the images. If detecting enough similar patches, the images can be identified as a rainy image. However, when there is no raindrop falling on the camera, it doesn't work. It also faced

misclassification by some objects such as lamps. Unfortunately, both of work focus on detecting rain, and don't put energy into different weather recognition and give a whole recognition result.

The contribution of this paper is to propose three groups of features according to images in different weather conditions, and give an algorithm derived from Real AdaBoost, which makes full use of category structure.

The rest of paper is organized as follows. In Section 2, we propose three groups of features by analyzing the images in different weather conditions. Real AdaBoost is introduced and the category structure is presented in Section 3. Section 4 shows perfect effect of our algorithm on our dataset captured by vision system in vehicle.

2 Feature Selection

For any pattern recognition problem, it is important to select proper features. Weather recognition from images is different from general image classification tasks. We always implement the image classification task by selecting interesting points as features or detecting the object emerging in the scene. It is impractical for our project because under different weather conditions there can be same objects and interesting points. Hence, applying the same kind of features as general image classification tasks is not proper. We propose typical features in low vision level by analyzing the property of images under the different weather conditions.

2.1 Histogram of Gradient Amplitude (HGA)

The images under different weather conditions take on different degrees of blur. In sunny days, it always makes the images sharper whereas blurred in rainy days. Especially when raindrop covers the camera, the images are always more blurred and the values of pixels in images are flatted. Gradient is a perfect tool of measuring sharpness [4][5]. Generally, the larger the gradient is, the more it is possible to be sunny. We compute the amplitude of gradient according to (1) and form a histogram of gradient amplitude.

$$M(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2}. \quad (1)$$

Fig.1 shows sunny and rainy images and their corresponding histograms of gradient amplitude. We find that the distribution of histogram is different. It is flatter for rainy images than sunny images. There are more low value pixels in rainy images and more high value pixel in sunny images.

2.2 Histogram of HSV Color Space (HSV)

According to our observation, brightness value is high in sunny images, and low in rainy images. There are more vivid pixels under the condition of sunny days and in

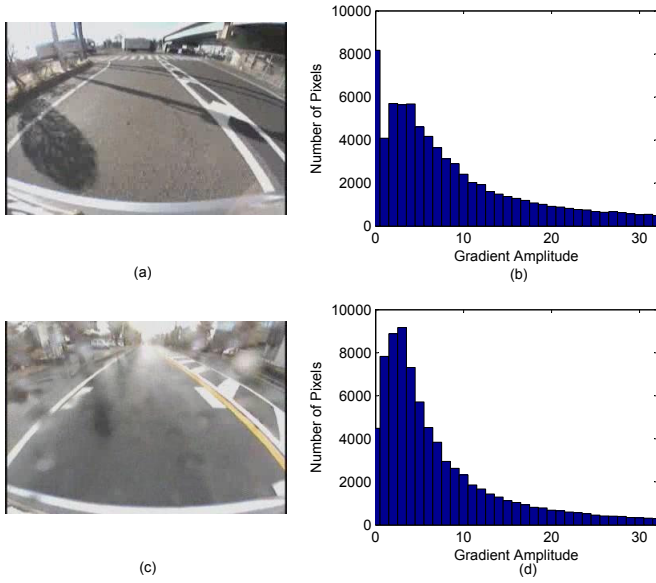


Fig. 1. (a) A sunny image; (b) Histogram of gradient amplitude of (a); (c) A rainy image; (d) Histogram of gradient amplitude of (c)



Fig. 2. ROI is surrounded by the red rectangle and the selected points is denoted by the red dots

contrast in rainy days [6]. It is also corresponding to human knowledge. Hence, HSV color space is thought to be a good measurement tool for weather classification. We convert the image into the HSV color space, portion the hue, saturation, value into 2,3,5 bins separately, and form a histogram as a group of feature.

2.3 Road Information (Road)

Not every patch of image contains the discriminative information for classification. Some areas of image contain more discriminative information than others and we often call it Region of Interest (ROI) [7]. In images captured by camera in vehicles, the road surface is always distinctive to human eyes, and we choose the central area as ROI. 10 points are selected in the road area, and the mean of gray value in 11×11 panes which are centered at the selected points is calculated. The ten values form a vector as a group of feature. Fig 2 shows the ROI which is surround by the red rectangle and the selected points are denoted by the red dots. ROI area is discriminative, but can not include all information in the images. Therefore, combining the global features and local features are our choice.

3 Recognition Algorithm

In this section, we describe our algorithm in detail. Real AdaBoost is introduced briefly in the first part. The category structure in weather recognition is proposed in part 2.

3.1 Real AdaBoost

AdaBoost is a powerful algorithm in pattern recognition and is widely used in the past ten years. Different from Support Vector Machine, Artificial Neural Network and Nearest Neighbor, AdaBoost combines many weak learners and forms a strong classifier. It has high accuracy and is resistant to overfitting. We choose Real AdaBoost [8][9] as our basic algorithm, which is shown to be better than discrete AdaBoost in most situations. We reproduce the algorithm procedure as follows. Suppose $S = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_N, y_N)\}$ as sample space, where $x \in X$ are feature vectors and $y \in \{-1, +1\}$ are labels. $D(i) = 1/N, i = 1, 2, \dots, N$ is the initial distribution.

For $t = 1 : T$

Step 1: Run a CART (Classification and regression tree) and get a best weak learner, and divide the sample space X into $X_1, X_2 \dots X_m$.

Step 2: Under $D(i)$, compute

$$\begin{aligned}
 p_l^j &= P(x_i \in X_j, y = l) \\
 &= \sum_{i: x_i \in X_j \wedge y_i = l} D_t(i) \quad l = \pm 1.
 \end{aligned}
 \tag{2}$$

Step 3: Set weak learner $\forall x \in X_j, h_t(x) = \frac{1}{2} \ln \left(\frac{p_{+1}^j + \epsilon}{p_{-1}^j + \epsilon} \right)$, where $j = 1, 2, \dots, m$, ϵ is a small positive number.

Step 4: Refresh the sample weights $D_{t+1}(i) = \frac{D_t(i) \exp[-y_i h_t(x_i)]}{Z_t}$, where Z_t is a normalized constant.

The final strong classifier is

$$H(x) = \text{sign} \left[\sum_{t=1}^T h_t(x) \right]. \quad (3)$$

3.2 Category Structure in Weather Recognition

In general image classification issues, there is no logical relation between classes and all classes are thought to be independent. As shown in Fig 3(a), the data of three classes overlap each other and we have to use three two-class classifiers to carry out the task. However, if the data of the extracted features from different classes distribute along a one-dimension manifold, i.e., a curve and the boundaries of these classifiers are approximately parallel in feature space as shown in Fig 3(b), we can think there exists a special category structure and it will be helpful for developing some simple algorithms for multi-class tasks.

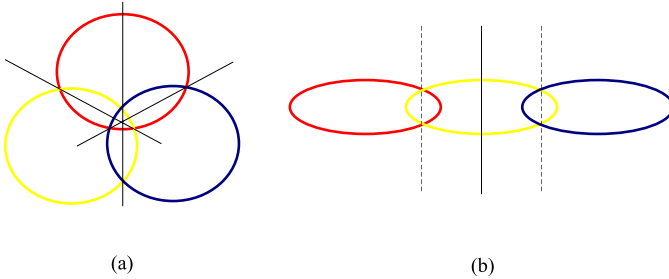


Fig. 3. Three different color circles or ellipses represent three different class data distribution

Our weather recognition issue under the extracted features is approximately agreed with the situation in Fig 3(b). The special category structure suggests that when a sample is misclassified into a rainy sample by Sunny-Rainy classifier, it has a higher probability to be a cloudy sample while not a sunny sample. Likewise, if a sample is misclassified into a sunny sample, it is more likely to be a cloudy one. That inspires us to devise an algorithm based on this category structure. We describe the algorithm as follows, which is shown in Fig 4.

Begin:

Step 1: Train the Sunny-Rainy, Rainy-Cloudy, Cloudy-Rainy classifiers.

Step 2: Input the testing samples into Sunny-Rainy classifier, and get the temporary label L_{temp} .

Step 3: If L_{temp} is sunny, put the sample into the Sunny-Cloudy classifier and give the final label L_{final} which belongs to sunny or cloudy. If L_{temp} is rainy, put the sample into the Cloudy-Rainy classifier and give the final label L_{final} which belongs to cloudy or rainy.

End.

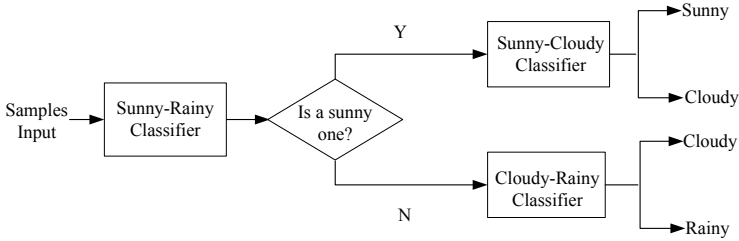


Fig. 4. The procedure of our algorithm based on category structure

One-vs-all makes the most basic algorithms suitable for multiple class problems. Suppose there are altogether K categories. One-vs-all need train K two-class classifiers and samples should also be tested in K classifiers. If the problem has the category structure what we have analyzed, our algorithm is much simpler than one-vs-all, because samples are only tested in $\lceil \log_2 K \rceil$ classifiers. Hence, our algorithm is faster and it also has approximately the same recognition rate as one-vs-all. The larger K is, the more efficient our algorithm is. For example, when $K = 8$, each sample is tested in 8 classifiers by using one-vs-all, while our algorithm only need to test 3 times in the case of finding a similar category structure. Our algorithm is easy to be extended to more weather categories in future.

4 Experiments

We applied the above features and algorithms to weather recognition problem. We try to learn three concepts (Sunny, Cloudy, Rainy) and 2496 images are collected as database which comprises training set and testing set. All the images are acquired from videos captured by vision system in vehicle. Different scenes (see from Fig 5) such as city street, highway, overpass are included in our database. The number of images in each category are shown in Table 1. The training set is randomly selected from database. Each experiment is repeated 20 times and the average result is reported. The number of nodes in CART weak learner is set 3, and the number of iteration in AdaBoost is 200. We run our experiment in C++ code on a CPU of AMD Sempron 3200+ with 1.5G RAM.

First we evaluate the effect of the three groups of feature. The result is shown in Fig 6. HSV is the best feature, which can be as high as even exceeding 90%. HGA is also perfect and able to reach 80% at its peak. Road performs not better

Table 1. Training and Testing set for our experiments

Category	Sunny	Cloudy	Rainy
$N_{training}$	400	400	400
$N_{testing}$	372	272	652

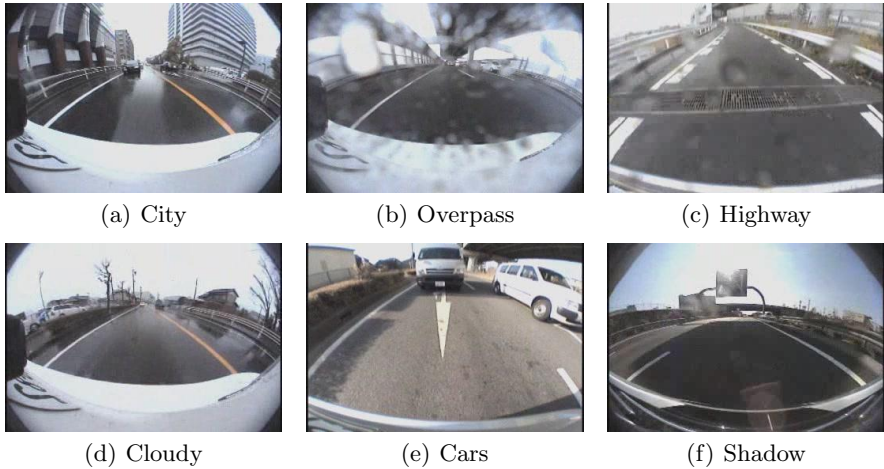


Fig. 5. Scenes included in our database

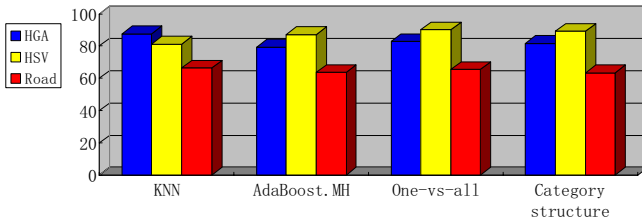


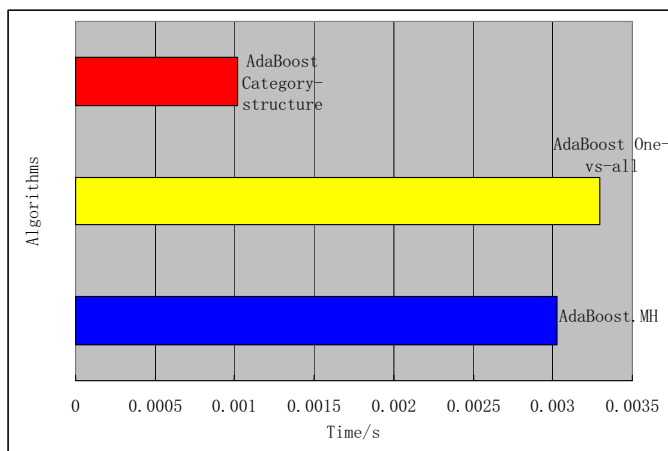
Fig. 6. Feature evaluation in different algorithms

than the other two because it is a group of local feature. Local feature do not include all information of images but it emphasizes the ROI and can serve as a good source for global features.

The result of algorithm evaluation is presented in Table 2. We introduce three other algorithms as comparison. K-Nearest Neighbor (KNN) is a general algorithm in pattern recognition [10]. Its basic idea is to find the K training samples which are nearest to the testing sample and give the most frequent label in K to the testing sample. AdaBoost.MH is proposed in [8], which can be directly used for multi-classification. AdaBoost with category structure is effective and its accuracy is higher than AdaBoost.MH and KNN, and only a little lower than AdaBoost with one-vs-all. However, our algorithm is obviously faster than other algorithms. From Fig. 7, AdaBoost with category structure costs only 1/3 of processing time of AdaBoost with one-vs-all and AdaBoost.MH. The time cost of KNN is not plotted, because its processing time is nearly 70 times of our algorithm, which would make the difference of the other three insignificant. All the processing time does not include the time of feature computing.

Table 2. Evaluation of algorithms

Category	Accuracy	Sunny Accuracy	Cloudy Accuracy	Rainy Accuracy
KNN	89.10%	93.51%	83.91%	89.61%
AdaBoost.MH	89.29%	94.22%	83.82%	89.82%
AdaBoost (one-vs- all)	92.15%	95.76%	88.81%	91.88%
AdaBoost (category structure)	91.92%	96.00%	89.35%	90.41%

**Fig. 7.** Time-consuming in different algorithms**Table 3.** The confusion matrix of AdaBoost with category structure result

Category	Sunny	Cloudy	Rainy
Sunny	0.9600	0.0345	0.0055
Cloudy	0.0249	0.8935	0.0816
Rainy	0.0037	0.0921	0.9041

From Table 3, the confusion matrix verifies our idea about category structure. The probability of sunny images misclassified into cloudy category is 0.0345, which is more than six times of sunny images misclassified into rainy category. Similarly, the rainy images is more likely to be misclassified into cloudy than sunny category.

5 Summary

We propose an algorithm based on AdaBoost for weather recognition, which uses HGA, HSV, Road three groups of features. The features are proved to be effective by our experiments. The category structure is introduced into our algorithm, which is combined with Real AdaBoost. The algorithm is shown to be efficient and effective by our experiments.

A limitation of our system is that it does not suit for some scenes which are also hard to judge by humans. When vehicles pass through a tunnel or a large shadow, it is hard to discriminate what kind of weather is. In future, we will utilize video processing technology to address this hard issue, in order to satisfy the practical application. Future research can focus on extending our three category weather recognition problem to more categories, such as foggy.

References

1. Garg, K., Nayar, S.K.: Detection and Removal of Rain from Videos. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 528–535. IEEE Press, New York (2004)
2. Kurihata, H., Takahashi, T., Ide, I., Mekada, Y., Murase, H., Tamatsu, Y., Miyahara, T.: Rainy Weather Recognition from In-Vehicle Camera Images for Driver Assistance. In: 2005 IEEE Intelligent Vehicles Symposium, pp. 205–210. IEEE Press, New York (2005)
3. Kurihata, H., Takahashi, T., Mekada, Y., Ichiro, I., Murase, H., Tamatsu, Y., Miyahara, T.: Raindrop Detection from In-Vehicle Video Camera Images for Rainfall Judgment. In: The First International Conference on Innovative Computing, Information and Control, pp. 544–547. IEEE Press, New York (2006)
4. Hsu, P., Chen, B.Y.: Blurred Image Detection and Classification. In: Satoh, S., Nack, F., Etoh, M. (eds.) MMM 2008. LNCS, vol. 4903, pp. 277–286. Springer, Heidelberg (2008)
5. Ng, K.C., Poo, A.N., Ang, M.H.: Practical Issues in Pixel-based Autofocusing for Machine Vision. In: Proceedings of the 2001 IEEE International Conference on Robotics and Automation, pp. 2791–2796. IEEE Press, New York (2001)
6. Renting, L., Zhaorong, L., Jiaya, J.: Image Partial Blur Detection and Classification. In: Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1–8. IEEE Press, New York (2008)
7. Bosch, A., Zisserman, A., Munoz, X.: Image Classification using Random Forests and Ferns. In: Proceedings of 2007 IEEE International Conference on Computer Vision, pp. 1–8. IEEE Press, New York (2007)
8. Schapire, R.E., Singer, Y.: Improved Boosting Algorithms Using Confidence-rated Predictions. *Machine Learning* 37, 297–336 (1999)
9. Friedman, J., Hastie, T., Tibshirani, R.: Additive Logistic Regression: A Statistical View of Boosting. *Annals of Statistics* 28, 337–374 (2000)
10. Duda, R.O., Hart, P.E., Stork, D.G.: *Pattern Classification*, 2nd edn. Wiley-Interscience, New York (2000)

Selecting Regions of Interest for the Diagnosis of Alzheimer Using Brain SPECT Images

Diego Salas-Gonzalez¹, Juan M. Górriz¹, Javier Ramírez¹, Ignacio Álvarez¹,
Míriam López¹, Fermín Segovia¹, and Carlos G. Puntonet²

¹ Dept. of Signal Theory, Networking and Communications, University of Granada,
Granada 18071, Spain

{dsalas,gorriz,javierrp}@ugr.es

² Dept. of Computer Architecture and Computer Technology, University of Granada,
Granada 18071, Spain
carlos@atc.ugr.es

Abstract. This paper presents a computer-aided diagnosis technique for improving the accuracy of the early diagnosis of the Alzheimer type dementia. The proposed methodology is based on the selection of those voxels which present a greater difference between normals and Alzheimer's type dementia patients. The mean value of the intensities of the selected voxels are used as features for different classifiers. The proposed methodology reaches an accuracy of 89 % in the classification task.

Keywords: Automatic Computer Aided Diagnosis, Classification, SPECT imaging, Alzheimer's disease.

1 Introduction

Nowadays, Single Photon Emission Computed Tomography (SPECT) is a widely tool in biomedical research and clinical medicine. SPECT imaging produces a mapping of physiological functions contrary to other imaging modalities which produce images of anatomical structures. SPECT provides three dimensional maps of a pharmaceutical labelled with a gamma ray emitting radionuclide. The distribution of radionuclide concentrations are estimated from a set of projectional images acquired at many different angles around the patient.

Single Photon Emission Computed Tomography imaging techniques employ radioisotopes which decay emitting predominantly a single gamma photon. When the nucleus of a radioisotope disintegrates, a gamma photon is emitted with a random direction which is uniformly distributed in the sphere surrounding the nucleus. If the photon is unimpeded by a collision with electrons or other particle within the body, its trajectory will be a straight line. A physical collimator is required to discriminate the direction of the ray by a photon detector external to the patient.

Brain SPECT imaging has become an important diagnostic and research tool in nuclear medicine. The use of brain imaging as a diagnostic tool in neurodegenerative diseases such as Alzheimer type disease (ATD) has been discussed

extensively [1, 2, 3]. Clinicians usually evaluate these images via visual inspection. Statistical classification methods have not been widely used for this task, possibly due to the fact that these images represent large amounts of data and most imaging studies have relatively few subjects (generally < 100). Despite of that, some works have been published recently [4, 5, 6, 7].

We study the overall difference between SPECT images of normal subjects and images from Alzheimer type disease patients. The set of voxels which presents greater differences between both categories are used as features for several classifiers. The proposed methodology allows us to obtain a very high accuracy in the classification of images (up to 89%).

This work is organised as follows: in Section 2 the different classifiers used in this work are presented; in Section 3, the SPECT image acquisition and pre-processing steps are explained; in Section 4, a procedure to select the voxels which will be used in the classification task is explained; in Section 5, we summarize the classification performance obtained applying various classifiers to the selected voxels; lastly, in Section 6 the conclusions are drawn.

2 Classifiers

The images we work with belong to two different classes: normal and Alzheimer type dementia (ATD). The goal of the classification task is to separate a set of binary labelled training data consisting of N -dimensional patterns \mathbf{x}_i and class labels y_i :

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_l, y_l) \in (R^N \times \{\pm 1\}), \quad (1)$$

so that a classifier f is produced which maps an object \mathbf{x}_i to its classification label y_i . The classifier f will correctly classify new examples (\mathbf{x}, \mathbf{y}) .

There are several different procedures to build the classifier f . In short, we utilize the following classifiers in this work [8].

2.1 Linear Classifier

A linear classifier classifies items which have similar features into groups by making a classification decision based on the value of the linear combination of the features. We Fit a multivariate normal density to each group, with a pooled estimate of covariance.

2.2 Mahalanobis Distance

We use Mahalanobis distances with stratified covariance estimates. The Mahalanobis distance differs from Euclidean in that it takes into account the correlations of the data set and is scale-invariant.

2.3 K-Nearest Mean

An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its k nearest neighbors. k is a positive integer, typically small. If $k = 1$, then the object is simply assigned to the class of its nearest neighbor. We choose $k = 5$ in the experimental results.

2.4 Support Vector Machines with Linear Kernels

Linear discriminant functions define decision hypersurfaces or hyperplanes in a multidimensional feature space:

$$g(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0 = 0 \quad (2)$$

where \mathbf{w} is the weight vector and w_0 is the threshold. \mathbf{w} is orthogonal to the decision hyperplane. The goal is to find the unknown parameters $w_i, i = 1, \dots, N$ which define the decision hyperplane.

Let $\mathbf{x}_i, i = 1, 2, \dots, l$ be the feature vectors of the training set X . These belong to two different classes, ω_1 or ω_2 . If the classes are linearly separable, the objective is to design a hyperplane that classifies correctly all the training vectors. This hyperplane is not unique and it can be estimated maximizing the performance of the classifier, that is, the ability of the classifier to operate satisfactorily with new data. The maximal margin of separation between both classes is a useful design criterion. Since the distance from a point \mathbf{x} to the hyperplane is given by $z = |g(\mathbf{x})| / \|\mathbf{w}\|$, the optimization problem can be reduced to the maximization of the margin $2 / \|\mathbf{w}\|$ with constraints by scaling \mathbf{w} and w_0 so that the value of $g(\mathbf{x})$ is $+1$ for the nearest point in w_1 and -1 for the nearest point in w_2 . The constraints are the following:

$$\mathbf{w}^T \mathbf{x} + w_0 \geq 1, \forall \mathbf{x} \in \mathbf{w}_1 \quad (3)$$

$$\mathbf{w}^T \mathbf{x} + w_0 \leq -1, \forall \mathbf{x} \in \mathbf{w}_2, \quad (4)$$

or, equivalently, minimizing the cost function $J(\mathbf{w}) = 1/2 \|\mathbf{w}\|^2$ subject to:

$$y_i(\mathbf{w}^T \mathbf{x}_i + w_0) \geq 1, i = 1, 2, \dots, l. \quad (5)$$

3 SPECT Image Acquisition and Preprocessing

The patients were injected with a gamma emitting ^{99m}Tc -ECD radiopharmaceutical and the SPECT raw data was acquired by a three head gamma camera Picker Prism 3000. A total of 180 projections were taken for each patient with a 2-degree angular resolution. The images of the brain cross sections were reconstructed from the projection data using the filtered backprojection (FBP) algorithm in combination with a Butterworth noise removal filter [9].

The complexity of brain structures and the differences between brains of different subjects make necessary the normalization of the images with respect to a common template. This ensures that the voxels in different images refer to the same anatomical positions in the brain. In this work, the images have been normalized using a general affine model, with 12 parameters [10].

After the affine normalization, the resulting image is registered using a more complex non-rigid spatial transformation model. The deformations are parameterized by a linear combination of the lowest-frequency components of the three-dimensional cosine transform bases [11]. A small-deformation approach is used,

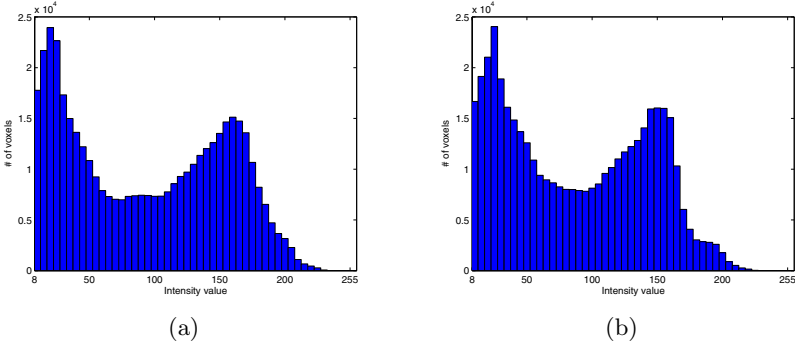


Fig. 1. Histogram of voxel intensities i with $I(i) > 8$. (a) Mean normal image \bar{I}_{NOR} . (b) Mean ATD image \bar{I}_{ATD} .

and regularization is by the bending energy of the displacement field. Then we normalize the intensities of the SPECT images with the maximum intensity, which is computed for each image individually by averaging over the 3% of the highest voxel intensities, similar as in [12].

4 Differences between Normal Subjects and Alzheimer's Disease Patients

Voxels which provide the greater difference between both groups (normal and ATD images) will be considered as training vectors of several classifiers. Therefore, this statistical study will allow us to select the discriminant regions of the brain to establish whether a given SPECT image belongs to a Normal or a ATD patient.

Figure 2 shows the distribution of the voxels for the mean normal images and mean ATD. It is easily seen that, if the whole image is considered, their distributions are quite similar. For this reason, it is convenient to select those voxels which present greater difference between intensity values for normal and ATD image.

Let the brain image set be I_1, I_2, \dots, I_N , where the number of images N is the sum of the images previously labelled as Normals (N_{NOR}) and Alzheimer type dementia (N_{ATD}) by expertises. Thus, the average Normal brain image of the dataset is defined as

$$\bar{I}_{NOR} = \frac{1}{N_{NOR}} \sum_{j \in NOR}^{N_{NOR}} I_j, \quad (6)$$

the average ATD can be calculated analogously:

$$\bar{I}_{ATD} = \frac{1}{N_{ATD}} \sum_{j \in ATD}^{N_{ATD}} I_j. \quad (7)$$

We study the overall difference between both subjects, normals and ATD. The image difference between the mean normal image and the mean ATD is depicted

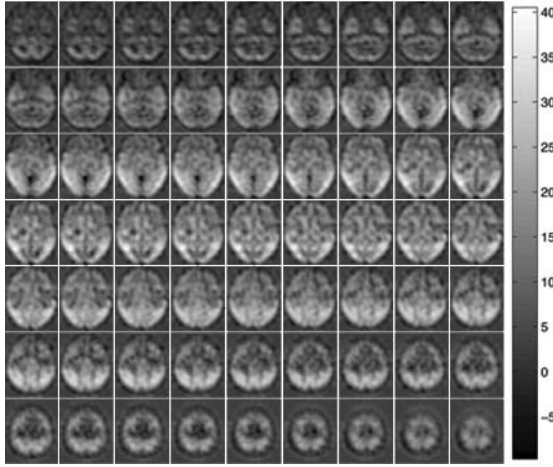


Fig. 2. Difference between mean Normal and mean ATD images

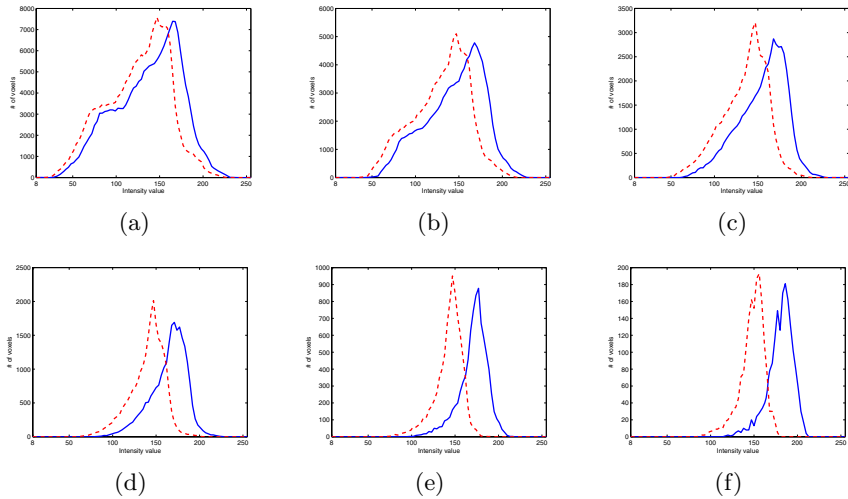


Fig. 3. Histogram with the intensity values of selected voxels with varied ε . Continuous line: mean normal image. Dotted line: mean ATD image. (a) $\varepsilon = 5$, (b) $\varepsilon = 10$, (c) $\varepsilon = 15$, (d) $\varepsilon = 20$, (e) $\varepsilon = 25$ and (f) $\varepsilon = 30$.

in Figure 2. This figure shows regions of hypoperfusion of the parieto-occipital cortices and bilateral parietotemporal hypoperfusion.

In the classification task, we consider those voxels i which present a difference greater than a given threshold ε .

$$i/\{|\bar{I}_{NOR}(i) - \bar{I}_{ATD}(i)| > \varepsilon\} \quad (8)$$

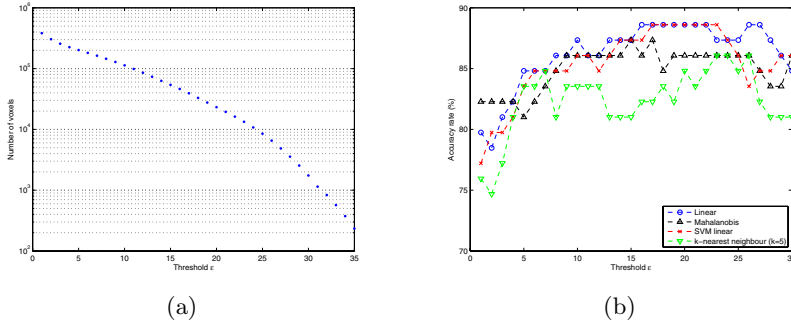


Fig. 4. (a) Number of selected voxels versus threshold value. (b) Accuracy rate versus threshold values ε for 4 different classifiers.

Figure 3 shows the distribution of the voxels i with $\bar{I}_{NOR}(i) - \bar{I}_{ATD}(i)$ greater than six different threshold values ε . It is easily seen that the difference between the histogram of the mean normal and ATD image increases concomitantly with the threshold value ε .

In Figure 4(a), the number of selected voxels versus the threshold value ε is plotted. Obviously, the number of voxels selected decrease with the value of ε . The amount of voxels selected by the condition (8) when ε is small is rather large to be considered as training vector in the classifiers. For this reason, we use the mean value of the intensities of the voxels which fulfill the condition $|\bar{I}_{NOR}(i) - \bar{I}_{ATD}(i)| > \varepsilon$.

5 Experimental Results

The performance of the classification is tested on a set of 79 real SPECT images (41 normals and 38 ATD) of a current study using the leave one-out method: the classifier is trained with all but one images of the database. The remaining image, which is not used to define the classifier, is then categorized. In that way, all SPECT images are classified and the success rate is computed from the number of correctly classified subjects.

The classification performance obtained with the different classifiers outlined in the preceding section versus the threshold value is shown in Fig. 4(b). We observe that the linear classifier and SVM with linear kernel perform similarly although the linear classifier performs better for higher threshold values. In general, if the value of the threshold increases, the images are classified more accurately. The best classification results are obtained with support vector machine and linear classifier with a threshold value of 16-23.

6 Conclusion

In this work, a straightforward criterion to select a set of discriminant voxels for the classification of SPECT brain images is presented. After normalisation of the

brain images, the set of voxels which presents greater overall difference between normals and Alzheimer type dementia images are selected. The mean value of the selected voxels are used as features to different classifiers. The proposed methodology reaches an accuracy of 89 % in the classification. Furthermore, it allows us to classify the brain images in normal and affected subjects in an automatic manner, with no prior knowledge about the Alzheimer disease.

Acknowledgment

This work was partly supported by the Spanish Government under the PETRI DENCLASES (PET2006-0253), TEC2008-02113, NAPOLEON (TEC2007-68030- C02-01) projects and the Consejera de Innovacin, Ciencia y Empresa (Junta de Andaluca, Spain) under the Excellence Project (TIC-02566).

References

- [1] Holman, B.L., Johnson, K.A., Gerada, B., Carvalho, P.A., Satlin, A.: The Scintigraphic Appearance of Alzheimer's Disease: A Prospective Study Using Technetium-99m-HMPAO SPECT. *Journal of Nuclear Medicine* 33(2), 181–185 (1992)
- [2] Jagust, W., Thisted, R., Devous, M.D., Heertum, R.V., Mayberg, H., Jobst, K., Smith, A.D., Borys, N.: Spect Perfusion Imaging in the Diagnosis of Alzheimer's Disease: A Clinical-pathologic Study. *Neurology* 56, 950–956 (2001)
- [3] McNeill, R., Sare, G.M., Manoharan, M., Testa, H.J., Mann, D.M.A., Neary, D., Snowden, J.S., Varma, A.R.: Accuracy of Single-photon Emission Computed Tomography in Differentiating Frontotemporal Dementia from Alzheimer's Disease. *J. Neurol. Neurosurg. Psychiatry* 78(4), 350–355 (2007)
- [4] Ramírez, J., Górriz, J.M., Romero, A., Lassl, A., Salas-Gonzalez, D., López, M., Gómez-Río, M., Rodríguez, A.: Computer Aided Diagnosis of Alzheimer Type Dementia Combining Support Vector Machines and Discriminant Set of Features. *Information Sciences* (accepted) (2008)
- [5] Fung, G., Stoeckel, J.: SVM Feature Selection for Classification of SPECT Images of Alzheimer's Disease Using Spatial Information. *Knowledge and Information Systems* 11(2), 243–258 (2007)
- [6] Górriz, J.M., Ramírez, J., Lassl, A., Salas-Gonzalez, D., Lang, E.W., Puntonet, C.G., Álvarez, I., López, M., Gómez-Río, M.: Automatic Computer Aided Diagnosis Tool Using Component-based SVM. In: *Medical Imaging Conference, Dresden. IEEE, Los Alamitos* (2008)
- [7] Lassl, A., Górriz, J.M., Ramírez, J., Salas-Gonzalez, D., Puntonet, C.G., Lang, E.W.: Clustering Approach for the Classification of Spect Images. In: *Medical Imaging Conference, Dresden. IEEE, Los Alamitos* (2008)
- [8] Krzanowski, W.J. (ed.): *Principles of Multivariate Analysis: A User's Perspective*. Oxford University Press, Inc., New York (1988)
- [9] Ramírez, J., Górriz, J.M., Gómez-Río, M., Romero, A., Chaves, R., Lassl, A., Rodríguez, A., Puntonet, C.G., Theis, F., Lang, E.: Effective Emission Tomography Image Reconstruction Algorithms for Spect Data. In: Bubak, M., van Albada, G.D., Dongarra, J., Sloot, P.M.A. (eds.) *ICCS 2008, Part I. LNCS, vol. 5101*, pp. 741–748. Springer, Heidelberg (2008)

- [10] Salas-Gonzalez, D., Górriz, J.M., Ramírez, J., Lassi, A., Puntónet, C.G.: Improved Gauss-newton Optimization Methods in Affine Registration of Spect Brain Images. *IET Electronics Letters* 44(22), 1291–1292 (2008)
- [11] Ashburner, J., Friston, K.J.: Nonlinear Spatial Normalization Using Basis Functions. *Human Brain Mapping* 7(4), 254–266 (1999)
- [12] Saxena, P., Pavel, D.G., Quintana, J.C., Horwitz, B.: An automatic threshold-based scaling method for enhancing the usefulness of Tc-HMPAO SPECT in the diagnosis of Alzheimer’s disease. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) *MICCAI 1998*. LNCS, vol. 1496, pp. 623–630. Springer, Heidelberg (1998)

Face Image Recognition Combining Holistic and Local Features

Chen Pan¹ and Feilong Cao²

¹College of Information Engineering, China Jiliang University,
Hangzhou 310018, China

²College of Science, China Jiliang University,
Hangzhou 310018, China
Pc916@cjlw.edu.cn

Abstract. This paper introduces a method using the holistic and the local features for face image recognition. The holistic feature is extracted from spatial domain by 2DPCA and the local feature is taken from 2D-DCT-frequency domain by 2DNMF, respectively. 2D-DCT coefficients form the different frequency components and get energy concentrate at the same time, which may be suitable to preserve some useful puny features often ignored in global method. And it may avoid the correlation between global and local features and offer complementary frequency information to spatial one. Finally, LSSVM regression is used to weight the mixed feature vectors and classify images. Experimental results have demonstrated the validity of the new method, which outperforms the conventional 2D-based PCA and NMF methods on ORL and JAFFE face databases.

Keywords: Feature extraction, Face recognition, 2DPCA, 2DNMF, LSSVM.

1 Introduction

Face recognition has been an active research area of computer vision and pattern recognition for decades. Many face recognition methods have been proposed to date, such as PCA[1], ICA[2], neural network[3], kernel methods[4], SVM[5], ensemble techniques[6], NMF[7] and NTF[8]. These methods could be roughly classified into three categories[9], according to the type of features used by various methods, i.e., holistic feature-based, local feature-based and hybrid feature-based methods. In the first category, the most often used method is the eigenface (PCA), which identify a image using the global feature of the image. While in the second category, the most widely used algorithm is the elastic bunch graph matching[10], local appearance-based methods[11], which use the local facial features for recognition. The third category is hybrid methods that are those approaches using both holistic and local features. The third category methods may have the potential to offer better performance than individual holistic or local methods, since more comprehensive information could be utilized. The key factors that influence the performance of hybrid methods include how to determine which features should be combined and

how to combine, so as to preserve their advantages and avert their disadvantages at the same time. These problems have close relationship with the multiple classifier system[12] and ensemble learning[13] in the field of machine learning. Unfortunately, even in these fields, these problems remain unsolved.

In fact, local features and global features have quite different properties and can hopefully offer complementary information about the classification task. In [9] summarizes qualitatively the difference between the two types of features and points that local features and global ones are separately sensitive to different variation factors. For instance, local feature is not sensitive to occlusion, but holistic feature is. While expression changes have more impact on holistic feature. For these observations, hybrid methods that use both holistic and local information for recognition may be an effective way to reduce the complexity of classifiers and improve their generalization capability.

Despite the potential advantages, the work in the third category is still relatively few, possibly due to the difficulties mentioned above. We observed that the most hybrid methods, such as flexible appearance models[14], hybrid LFA[11], extract features only in spatial domain. The spatial and frequency features are seldom used together at features level in recognition task.

In this paper, we presented a hybrid method combine holistic and local features to recognize a face. Different from the traditional methods, the new method adopts 2DPCA to extract holistic features from spatial domain and uses 2DNMF to extract local ones from 2D-DCT domain. Then every face image can be represented by a mixed feature vector and classified by LSSVM regression. The structure of the new method is shown in Fig.1.

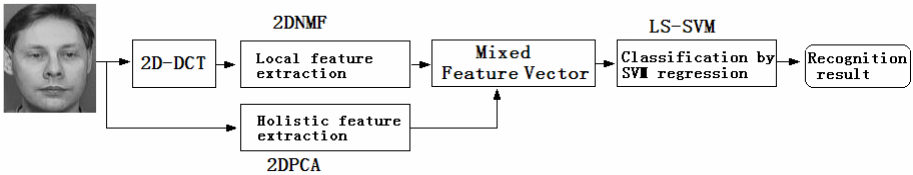


Fig. 1. The structure of the proposed method

The remainder of this paper is organized as follows: In Section 2, why the 2DPCA and 2DNMF are adopted here and how to use them for feature extraction are described. Multi-class Classification using LSSVM regression is developed in Section 3. In Section 4, experimental results are presented for the ORL and JAFFE to demonstrate the effectiveness of the new method. Finally, conclusions are presented in Section 5.

2 Image Representations Using the Holistic and the Local Features

Image feature extraction in template-based method executes dimensionality reduction actually. In this section, PCA-based method is adopted to extract the holistic feature, because PCA has average performance and less computation cost than some related

methods, such as KPCA. Recently, the dimensionality reduction solutions trend to represent images as matrices rather than vectors. The 2D-based method has many advantages over conventional 1D-based one. First, it is simpler and more straightforward to use. Second, it is better at recognition accuracy in most conditions. Third, it is computationally more efficient, and can speed up image feature extraction significantly [15].

Here, we adopted 2D-based strategy followed from [16]. In which 2D-based algorithm takes into account the spatial correlation of the image pixels within a localized neighborhood. Two linear transforms are applied to both the left- and the right-hand sides of the input image matrices. Thus, projections in both modes are calculated and the dimensionality reduction results can be obtained.

2.1 Using 2DPCA Extracts the Holistic Feature

Suppose that there are M training object images, denoted by $m \times n$ matrices A_k ($k = 1, 2, \dots, M$). Two type of the covariance matrixes can be define as

$$G_{row} = \frac{1}{M} \sum_{k=1}^M (A_k - \bar{A})^T (A_k - \bar{A}) \quad (1)$$

$$G_{col} = \frac{1}{M} \sum_{k=1}^M (A_k - \bar{A}) (A_k - \bar{A})^T \quad (2)$$

where $\bar{A} = \frac{1}{M} \sum_k A_k$ is the mean image. According to Eq.(1), the projective vectors

X_1, \dots, X_d can be obtained by computing the d eigenvectors corresponding to the d biggest eigenvalues of G_{row} . Since the size of G_{row} is only $n \times n$, computing its eigenvectors can be efficient. Similarly, according to Eq.(2), the projective vectors Y_1, \dots, Y_d can be obtained by computing eigenvectors of G_{col} .

Let $X = [X_1, \dots, X_d]$, $Y = [Y_1, \dots, Y_d]$ denote the left- and right-hand projective matrix respectively, projecting training image A_k s onto X and Y , yielding $d \times d$ feature matrices

$$C_k = Y^T A_k X \quad (3)$$

2.2 Using 2DNMF Extracts the Local Feature

Let us to see the Eq.(3), the left- and right-hand projective matrixes come from two covariance matrixes, it constructs two orthogonal basis matrixes to extract global features. From the view of matrix decomposition, if we find another way to construct those two basis matrixes, may be extract another feature from the image.

It is known that non-negative matrix factorization (NMF) is proposed to find the parts-based representation of non-negative data [7]. The key ingredient of NMF is the non-negativity constraints imposed on the two factors, so the constraints are compatible with the intuitive notion of combining parts to form a whole. Since a part-based

representation can naturally deal with partial occlusion and some illumination problems [7,17], it has received much attention recently.

Similarly to 2DPCA mentioned above, we present an improved 2DNMF algorithm not only represents the input image as a matrix, but also uses orthogonalized basis matrixes to detect local feature by learning in a lower dimensional subspace.

Assume that there are M training object images, denoted by $m \times n$ matrices A_k ($k = 1, 2, \dots, M$). Denote $B_k = A_k^T$, we can construct two training matrixes $V_{row} = [A_1, \dots, A_k]_{m \times nM}$ and $V_{col} = [B_1, \dots, B_k]_{n \times mM}$.

In order to reduce the dimensionality, NMF finds two non-negative matrix factors W and H such that

$$V_{row} \approx W_{m \times r} H_{r \times nM} \tag{4}$$

Here the r columns of W are called NMF bases (W is basis matrix), and the columns of H are its combining coefficients. Similarly,

$$V_{col} \approx W_{n \times r} H_{r \times mM} \tag{5}$$

The rank r of the factorization is usually chosen such that $(n + mM)r < nmM$, and hence dimensionality reduction is achieved.

Classical NMF is an iterative algorithm and the bases learned via NMF are not orthogonal. Liu[18] presented an orthonormalization strategy that first orthonormalize the bases, and then use the projections based on the orthonormalized bases for classification. So that the iterative operation of NMF in test could become a matrix operation that leads in less time cost and more precise in computation. We set

$$X_{m \times r} = orth(W_{m \times r}) \tag{6}$$

$$Y_{n \times r} = orth(W_{n \times r}) \tag{7}$$

Projecting training image A_k s onto X and Y , yielding $r \times r$ feature matrices D_k .

$$D_k = Y^T A_k X \tag{8}$$

The mixed feature combining holistic and local features could be assembled and denoted by

$$F_{(d^2+r^2) \times 1} = [C_k(\cdot), D_k(\cdot)]^T \tag{9}$$

In order to achieve more complementary components and avoid disadvantages such as correlation between local and global features, we transform image into 2D-DCT domain for extracting the local feature. So the A_k and B_k in this section are 2D-DCT coefficient matrixes, which form the different frequency components and get energy concentrate at the same time, thus may offer local frequency information to spatial one.

There is a problem should be solved in practice. That is the DCT coefficients may be negative, yet NMF requires non-negative data to decomposition. While we observed that the AC coefficients of DCT are derived from the intensity change of the

local pixels, i.e. from bright to dark or reverse. The absolute value of DCT coefficients of an image seem to be equal to that of its negative image, their DC component is nearly same and AC components only differ from in symbol. So we can represent an image using the absolute value of its DCT coefficients since it can represent both the image and its negative one. See Fig.2 shows.

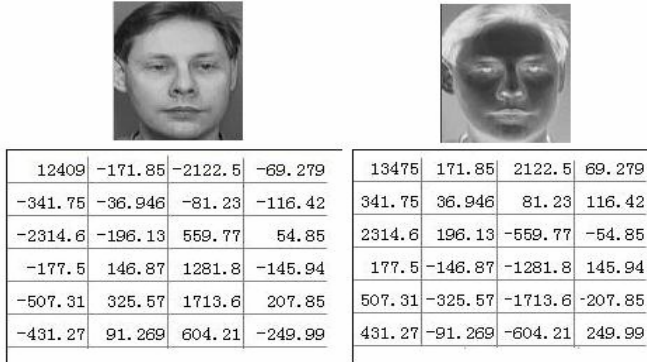


Fig. 2. An image and its negative image corresponding to their DCT coefficients (partly)

3 Multi-class Classification Using LS-SVM Regression

Support Vector Machines (SVM)[5] is a powerful methodology for solving problems in nonlinear classification and function estimation which has also led to many other recent developments in kernel based methods in general. Originally, it has been introduced within the context of statistical learning theory and structural risk minimization. In the methods one solves convex optimization problems, typically by quadratic programming. Least Squares Support Vector Machines (LS-SVM) are reformulations to the standard SVMs[19]. The cost function is a regularized least squares function with equality constraints, leading to linear Karush-Kuhn-Tucker systems. The solution can be found efficiently by iterative methods.

Traditional multi-classifier using SVM is usually constructed by one-vs-one or one-vs-all strategies. It may result in a blind area for classification in feature space, in which some samples could not be predicted. In order to avoid this problem we use SVM regression to do classification because the classification and regression are equivalent in nature.

Let the training set is $\{\mathbf{x}_i, y_i\}_{i=1}^l$, with input $\mathbf{x}_i \in R^d$, when $y_i = \{\pm 1\}$ means binary classification task, and $y_i \in R$ means regression one. While multi-class task often is $y_i = \{1, 2, \dots, l\}$, l denotes the label value of class. So multi-class classification task could be regard as a special case in regression if we denote the class label with different numerical value. An algorithm using LS-SVR for multi-classification (denoted by LS-SVRC) is presented bellow:

Step1: Set the class label value l_j in training sets.

Step2: Execute LS-SVM training to get a regression model.

Step3: Predict $f(\mathbf{x})$, the value of the test sample using the regression model;

Step4: Estimate the label j^* using formula $j^* = \arg \min \{|f(\mathbf{x}) - l_j|\}$.

By this way, we can get an exclusive label corresponding to pre-class. Furthermore, it may be a new way to achieving better result by specifying the class label actively according to the characteristic of every class.

4 Experimental Result

In this section, we test the proposed (hybrid-feature-based) method compared with the mentioned 2DPCA(holistic-feature-based method) and 2DNMF(local-feature-based method) on the ORL [20] and JAFFE [21] face database. All the experiments are carried out on PC with P4 1.7GHz CPU and 512MB memory in Matlab 7. The kernel type of LS-SVM is RBF. There are two parameters need to adjust according to the training data. One is the kernel parameter, the other is the regularization constant γ . Tuning those parameters may refer to [22].

4.1 ORL Database

The ORL database contains images from 40 individuals, each providing 10 different images. The first 5 images for each subject are used for training, and the rest for testing.

Firstly, the nearest neighbor (NN) classifier and Euclidean distance are used for classification, after extracting the features using the above mentioned methods. Table 1 shows the comparison of the top recognition accuracy of the four methods. Here the recognition accuracy is defined as the percentage of correctly recognized images in test set.

In Table.1, 2DPCA, 2DNMF and 2DPCA+2DNMF denote that features extracted only from spatial domain. 2DPCA+2DNMF(DCT) represents the hybrid method. Table.1 shows combining 2DPCA and 2DNMF method in spatial domain does not enhance the accuracy although the dimensionality of the mixed feature vector increased. The extended features may not be always benefit the aim of classification. Yet in this example, the extended features seem not deteriorate the recognition performance. While

Table 1. Comparison on top recognition accuracy using 1NN classifier and Euclidean distance, the corresponding dimensions of feature matrices, and the running time cost in test(second)

Method	Accuracy(%)	Dimension	Time(s)
2DPCA+NN	91.5	20*20	0.81
2DNMF+NN	90.5	50*50	0.89
2DPCA+2DNMF+NN	91.5	10*10+50*50	1.4
2DPCA+2DNMF(DCT)+NN	92.5	10*10+50*50	1.5

extracting the features from spatial and frequency domain can improve the recognition accuracy. The useful frequency information could complement to spatial one obviously. Because Euclidean distance is used here and the dimensionality of the local feature on the top accuracy is over decuple of that of the holistic one, that means less correlation between the local and the holistic features at least.

Secondly, the LS-SVRC classifier is used for classification. We compared four feature-extracting methods in Fig.3. In which, 2DPCA or 2DNMF denote that only 2DPCA or 2DNMF method used in spatial domain; 2DNMF(DCT) denotes that method used in DCT domain, and 2DPCA+2DNMF(DCT) means our hybrid method in this paper (whose dimensions is $(d*d+r*r)$).

From Fig.3, the SVM classifier enhances all the accuracy of the methods significantly, but the top one appears in our hybrid method. First reason is that SVM-based method overcomes the correlation between different features via weighting them in training[19]. Second, our hybrid feature extraction strategy also gives a forceful contribution. 2D-DCT coefficients form the different frequency components and get energy concentrate at the same time, which may be suitable to preserve some useful puny features via NMF, which often ignored in global method. So it may offer complementary frequency information to spatial one. Third, PCA and NMF with orthonormalized bases also is an effective way due to the orthonormalization also eliminates the interrelationship among features.

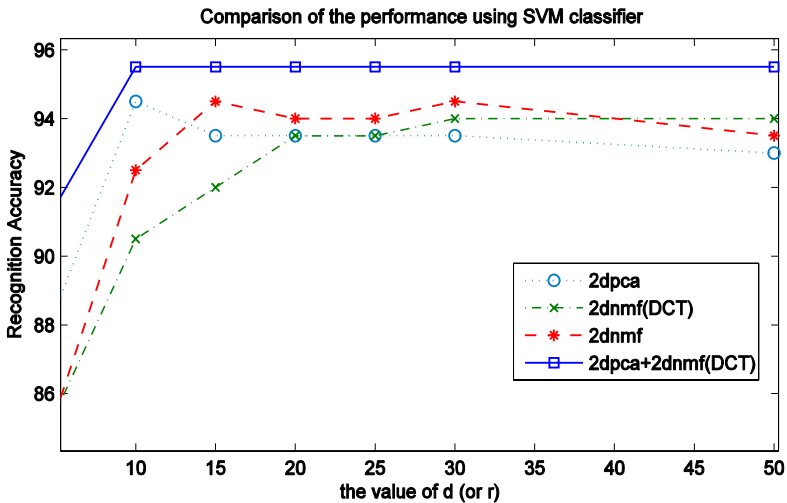


Fig. 3. Comparison of recognition accuracy according to their dimensions in ORL database. The dimensions of the hybrid method is $(d*d+r*r)$, that of the rest methods are $d*d$ or $r*r$.

4.2 JAFFE Database

The JAFFE database contains 213 images from 10 Japanese female models, each providing more than 20 different images with 7 facial expressions at least. We selected 10 images randomly for each subject for training, and the rest 10 images for testing. Similar as Fig.3, we compared the mentioned feature-extracting methods in Fig.4.

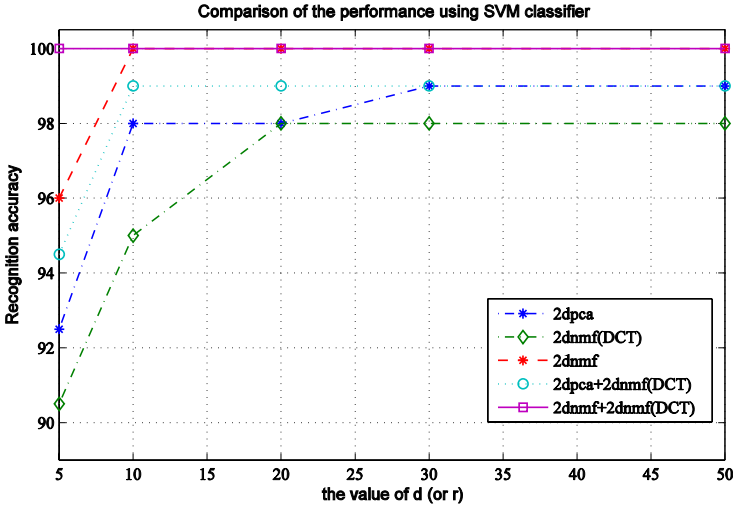


Fig. 4. Comparison of recognition accuracy in JAFFE database

From the Fig.4, the top accuracy appears in 2DNMF+2DNMF(DCT) method. The pure 2DNMF method also shows good performance whose rate has only a slight decline in low-dimensional. It illustrates that the part-based method is more suitable to deal with the change of expression in JAFFE database. Yet, the accuracy rate of 2DPCA+2DNMF(DCT) method is still at high position. The performance of the hybrid approaches is better than that of the individual 2DPCA and 2DNMF(DCT) methods in low-dimensional especially. It means the local frequency feature could benefit to image recognition whether holistic feature-based or local feature-based methods.

5 Conclusions

We introduced a method using the holistic and the local features for face image recognition. The holistic feature is extracted from spatial domain by 2DPCA and the local feature is taken from 2D-DCT-frequency domain by 2DNMF, respectively. 2DNMF is used to decompose the coefficient matrix of 2D-DCT could extract the different frequency components, which can offer complementary frequency information to spatial one and avoid the correlation between global and local features at same time. Finally, LSSVM regression is used to select the mixed feature vectors and classify images. In most cases, the new method outperforms the conventional 2D-based PCA and NMF methods. It has demonstrated that the presented method is effective.

Since many image and video data are naturally tensor objects, the future work is to use the tensor-factorization-based method such as MPCA[23] and NTF[8] instead of 2DPCA and 2DNMF respectively to realize the hybrid feature extraction, by which the proposed method may achieve more benefits.

Acknowledgments. This work was supported by the National Science Foundation of China under grant (No. 90818020 and 60873206). And Supported by the Natural Science Foundation of Zhejiang Province of China (No.Y7080235).

References

1. Turk, M.A., Pentland, A.P.: Face Recognition Using Eigenfaces. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 586–591 (1991)
2. Comon, P.: Independent Component Analysis, A New Concept? *Signal Process* 36(3), 287–314 (1994)
3. Bishop, C.M.: *Neural Network for Pattern Recognition*. Oxford University Press, New York (1995)
4. Lu, J., Plataniotis, K.N., Venetsanopoulos, A.N.: Face Recognition Using Kernel Direct Discriminant Analysis Algorithms. *IEEE Trans. on Neural Networks* 14(1), 117–126 (2003)
5. Vapnik, V.: *The Nature of Statistical Learning Theory*. Springer, New York (2000)
6. Pang, S., Kim, D., Bang, S.Y.: Membership Authentication in the Dynamic Group by Face Classification Using SVM Ensemble. *Pattern Recognition Letters* 24(1-3), 215–225 (2003)
7. Lee, D.D., Seung, H.S.: Learning the Parts of Objects by Non-negative Matrix Factorization. *Nature* 401, 788–791 (1999)
8. Shashua, A., Hazan, T.: Non-Negative Tensor Factorization with Applications to Statistics and Computer Vision. In: *Proceedings of the 22nd International Conference on Machine Learning*, Bonn, Germany (2005)
9. Tan, X.Y., Chen, S.C., Zhou, Z.H., Zhang, F.Y.: Face Recognition from A Single Image Per Person: A survey. *Pattern Recognition* 39, 1725–1745 (2006)
10. Wiskott, L., Fellous, J.M., Kruger, N., Malsburg, C.: Face Recognition by Elastic Bunch Graph Matching. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 19(7), 775–779 (1997)
11. Penev, P., Atick, J.: Local Feature Analysis: A General Statistical Theory for Object Representation. *Netw.: Comput. Neural Syst.* 7, 477–500 (1996)
12. Kittler, J., Hatef, M., Duin, R.P.W., Matas, J.: On Combining Classifiers. *IEEE Trans. Pattern Anal. Mach. Intell.* 20(3), 226–239 (1998)
13. Zhou, Z.H., Wu, J., Tang, W.: Ensembling Neural Networks: Many Could be Better Than All. *Artif. Intell.* 137(1-2), 239–263 (2003)
14. Lanitis, A., Taylor, C.J., Cootes, T.F.: Automatic Face Identification System Using Flexible Appearance Models. *Image Vision Comput.* 13, 393–401 (1995)
15. Yang, J., Zhang, D.Q., Yang, J.Y.: Two-dimensional PCA: A New Approach to Appearance-based Face Representation and Recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 26(1), 131–137 (2004)
16. Zhang, D.Q., Zhou, Z.H.: (2D)²PCA: 2-directional 2-dimensional PCA for Efficient Face Representation and Recognition. *Neurocomputing* 69(1-3), 224–231 (2005)
17. Zhang, D.Q., Chen, S.C., Zhou, Z.H.: Two-Dimensional Non-negative Matrix Factorization for Face Representation and Recognition. In: Zhao, W., Gong, S., Tang, X. (eds.) *AMFG 2005*. LNCS, vol. 3723, pp. 350–363. Springer, Heidelberg (2005)
18. Liu, W.X., Zheng, N.N.: Non-negative Matrix Factorization Based Methods for Object Recognition. *Pattern Recognition Letters* 25, 893–897 (2004)
19. Suykens, J.A.K., Vandewalle, J.: Least Squares Support Vector Machine Classifiers. *Neural Processing Letter* 9, 293–300 (1999)
20. ORL, <http://mambo.ucsc.edu/psl/olivetti.html>
21. JAFFE, <http://www.kasrl.org/jaffe.html>
22. Pelckmans, K., Suykens, J.A.K., Van Gestel, T., et al.: *LS-SVMLab Toolbox User's Guide version 1.5.*, <http://www.esat.kuleuven.ac.be/sista/lssvmlab/>
23. Lu, H.P., Konstantinos, N., Plataniotis, Venetsanopoulos, N.A.: MPCA: Multilinear Principal Component Analysis of Tensor Objects. *IEEE Trans. on Neural Networks* 19(1), 18–39 (2008)

3D Representative Face and Clustering Based Illumination Estimation for Face Recognition and Expression Recognition

Zheng Zhang^{1,3}, Zheng Zhao¹, and Gang Bai²

¹ College of Computer Science and Technology, Tianjin University, Tianjin 300072, China

² College of Information Technical Science, Nankai University, Tianjin 300071, China

³ Tianjin University of Technology, Tianjin 300191, China

aaron_boy_2000@hotmail.com, zhengzh@tju.edu.cn,
baigang@nankai.edu.cn

Abstract. Eliminating the negative effect caused by variant pose and illumination is a very critical problem for expression recognition. In this paper we propose a 3D representative face (RF) and clustering based method, which can estimate 13 illumination conditions under certain poses. First, all faces are adaptively categorized into 31 facial types by k-means clustering, so people with similar facial appearance are clustered together; Then the representative face of each cluster is generated. Finally we select the most discriminative features to train a group of SVM classifiers and get 96.88% estimation accuracy when estimating the test set with frontal view. Compared with other related works, ours does not rely on 3D reconstruction, and to get the generalization ability, we use our RF and clustering technique.

Key words: Illumination estimation, K-Means clustering, Representative face.

1 Introduction

In the last few years, with the rapid progress of Human-computer intelligent interaction (HCII), expression recognition has become a very active topic in machine vision community. In [1] we can see that for expression recognition, now the main challenge is the recognition under variant pose and illumination (PI).

There are many 3D reconstruction [2, 3] based former works in face recognition, which aim at eliminating the negative effect caused by variant PI. But there are two problems with reconstruction based method: first, not independent with the subjects' identity, we call it a generalization problem; second, its high computational complexity — $O(M \times N)$, where M is the number of illuminations and N is the number of 3D models in training set. In this paper we propose a novel illumination estimation method that can be done without 3D reconstruction, which solves the generalization problem by our representative face(RF) and clustering technique with complexity $O(C \times N)$, where C is the number of clusters, a constant.

We believe the estimation of PI need to be done in 2 steps and since it is more easier to get illumination-invariant descriptors, the 1st step should be the estimation

of pose [4]. In this article, we aim at the estimation of illumination conditions under certain poses. That is if the pose of an input image is known, our method can estimate the illumination conditions of that pose. We define 13 lamp-house positions to form 13 illumination classes, and test the estimation performance at a series of poses with the pan angle spanning from -60° to 60° and the tilt angle spanning from -40° to 40° . See fig. 1 for illustration.

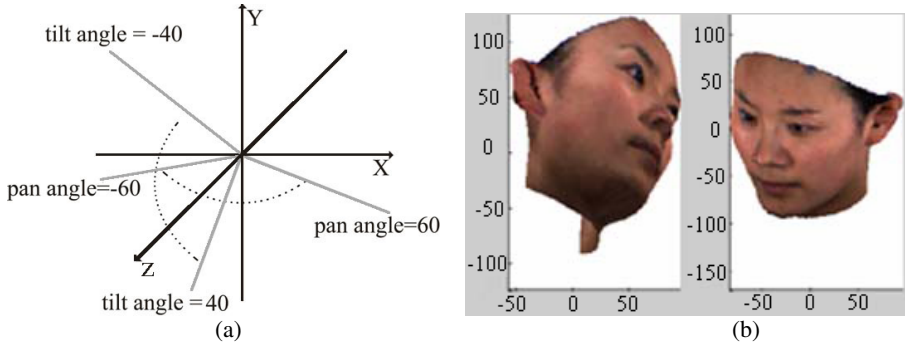


Fig. 1. (a) Tilt angle & pan angle in pose definition; (b) Two 3D models under certain poses

The rest of the paper is organized as follows. In the next section, we give an overview of our 3D representative face and clustering based framework. In section 3 we apply adaptive k-means clustering to solve the generalization problem. Section 4 presents the concept of 3D representative face, namely RF. In section 5 we introduce the features and the SVM classifier we employ in illumination estimation. In section 6 we show the experimental results. Conclusions are given in the last section.

2 Illumination Estimation: An Overview

The dataset we use is BJUT-3D Face Database [5]. To make our RF and clustering based method work well, we hope our train set contains as much types of facial appearance as possible. So we randomly select 320 subjects and their 3D face models are divided into 5 parts randomly, 256 models are used to generate the representative faces (RF) for training, the other models are kept to generate 64×13 2D test images with variant illuminations for testing the generalization ability of our method.

First, all 256 faces are categorized into 31 classes based on the positions of their fiducial points by adaptively k-means clustering, so people with the similar facial appearance are clustered together; Then the 3D average face of each cluster is generated to represent the facial appearance type of that cluster (so-called RF). By rotating all RFs to a certain pose, illuminating them with 13 illumination conditions, and projecting them to 2D we get all the 13-class 2D virtual lighting faces as our train set images; Finally we select the most discriminative rice features to train a group of SVM classifiers for a 13-class problem and get satisfactory estimation accuracy when estimating the test set with 64×13 samples.

3 Adaptive Facial 3D Structure K-Means Clustering

It is a person’s 3D facial structure that determines the appearance (intensity distribution) of his photo under variant illuminations, and facial organs such as eyes, nose and mouth is the main cause of 3D structure difference among different people. Though people’s facial appearances differ in thousands of ways, their facial structures can be classified into some main types according to the positions and shapes of their facial organs. By clustering all 256 3D face models according to the coordinates of their main facial organs, we actually cluster 256 face models into a number of facial structure types.

After detecting the four critical points in 3D models, we have all models’ nose tips aligned to a base point (x_0, y_0, z_0) . Now the coordinates’ of the rest three critical points are used as features for facial structure:

$$V = (x_1, y_1, z_1, x_2, y_2, z_2, x_3, y_3, z_3)^9$$

It is difficult to decide an appropriate cluster number for k-means clustering algorithm if we do not understand the inner-structure of the data well. Usually better choice of cluster number is crucial to the clustering result. In this paper, we adaptively get the cluster number between 15 and 35 following the max mean Silhouette value principle. When clustering we randomly select the cluster centers and repeat 5 times with different starting points in case of local minima. Finally we get the cluster number 31.

To get an idea of how well-separated the resulting clusters are, see Fig. 2.(b) for a silhouette plot. The silhouette plot displays a measure of how close each point in one cluster is to points in the neighboring clusters. This measure ranges from +1, indicating points that are very distant from neighboring clusters, through 0, indicating points that are not distinctly in one cluster or another, to -1, indicating points that are probably assigned to the wrong cluster.

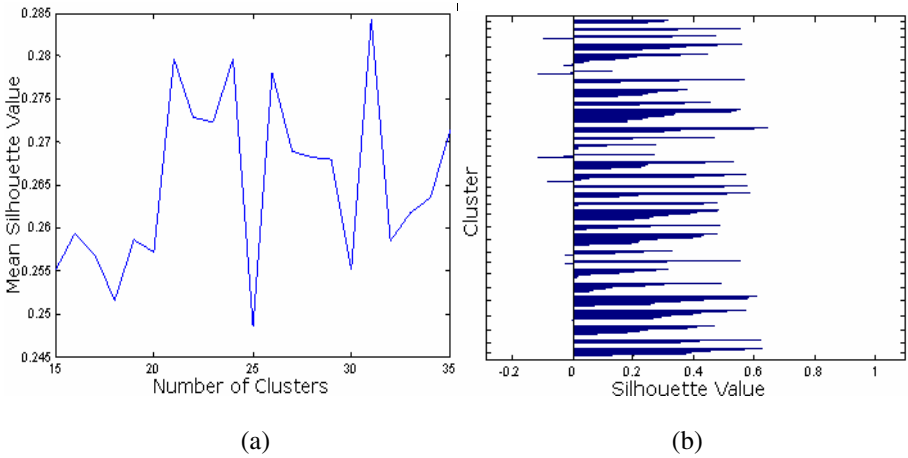


Fig. 2. (a) Choice of Clusters Numbers, (b) Silhouette Value for 31 Clusters

4 Generating Representative Face—RF

A 3D average face represents a kind of 3D stable structure hidden behind all individual faces who contribute in computing it. We believe that for the need of illumination estimation, it is stable and representative enough to approximate all individual faces belonging to its facial structure type. So we generate an average face [6, 7] as the representative face for each cluster to represent 31 types of facial structures. For illustration, Fig. 3 shows 6 RFs out of total 31.

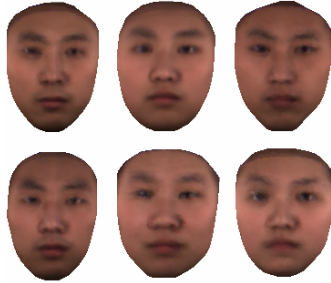


Fig. 3. Six RFs of six clusters representing six different facial structure types

5 Feature Selection and Classification

5.1 Generating Train Set

In our experiment, we define 13 kinds of illuminations. As shown in Fig. 4. To get train samples for our 13 illuminations estimation problem. First, we illuminate all 31 RFs, each with the 13 kinds of illuminations defined above. Then we project all 31×13 illuminated RFs to 2D to get training images for our 13-class problem. For each new test sample, we can always expect that there is a facial structure type it belongs to in our 31 RFs, and so each class has a training sample from that facial structure type, from which we get the generalization ability. This is the essential of our RF and clustering based method.

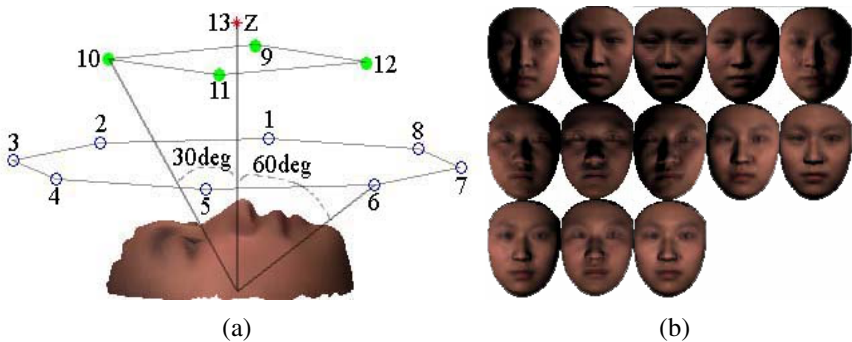


Fig. 4. (a) Positions of the 13 light-houses, (b) A representative face under 13 illuminations

5.2 Rice Feature

In a pattern recognition problem, it is of great significance to get the most discriminative feature for classification. In this paper, we select pixels which are most sensitive to illumination changes. We call it ‘Rice Feature’ because its shape is like the Chinese word ‘Rice’ – “米”.

As illustrated in Fig. 5, four real lines form the Chinese ‘Rice’ – “米”. We select 22 pixel-lines(4 real lines plus 18 dashed lines) of 4 directions from training images. By using rice features we can get comparable recognition accuracy with pixel-features extracted from the face region while the length of the feature vector is much more smaller.

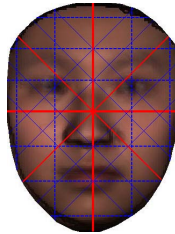


Fig. 5. Rice feature

5.3 Support Vector Machines

Unlike many traditional classifiers that aim at minimizing the Empirical Risk, SVM approaches the classification problem as an approximate implementation of the Structural Risk Minimization (SRM) induction principal, which may means better generalization ability.

5.4 Multi-class Classification

SVM, as explained above, is suitable only for binary classification, while our illumination estimation is a 13-class problem. However, there are many techniques that can extend SVM to handle a multi-class problem. In our experiment, we try two techniques: (1) “one-against-one” voting strategy [8]; (2) “one-against-one” eliminating strategy [9].

When using the “one-against-one” approach we get satisfactory results with both voting strategy and eliminating strategy, and the performance of eliminating strategy is a little higher (96.88% versus 96.75% when estimating illuminations with frontal view). In our opinion, this may because that with the voting strategy, in each binary classification the only information we can get is yes or no(+1 or -1), while with the eliminating strategy, a real value between -1 and +1 (yes if >0, no if <0) is given as the confidence of that classification. Though this brings no difference in a 2-class problem, more information is provided for multi-classification.

6 Experimental Results

In our experiments, illumination estimations under a series of poses from the pan angle spanning from -60° to 60° and the tilt angle spanning from -40° to 40° are tested. For each test we give two estimation results, one is for the 3D face models participating in the generation of the representative faces, we project these models with 13 illuminations to 2D to form 256×13 images, and we call it group-I, the other is for the 64×13 test images, we call it group-II. Some typical results are outlined below in table 1. Data are formatted as group-I / group-II. We omit the results when pan angle is -30 or -60 because of symmetry.

Table 1. The classification results under certain poses

Pan angle \ Tilt Angle	0	30	60
-40	92.22% / 91.59%	93.60% / 94.23%	92.01% / 90.38%
-20	94.35% / 92.67%	92.07% / 93.60%	93.90% / 92.07%
0	97.39% / 96.88%	96.94% / 96.51%	95.10% / 93.99%
20	90.53% / 89.54%	90.99% / 90.87%	88.43% / 88.46%
40	82.48% / 82.57%	80.38% / 81.01%	73.23% / 72.84%

7 Conclusions and Future Works

In this paper, a 3D representative face and clustering based illumination estimation framework is proposed. From the estimation results shown above, we can see that both accuracies of group-I and group-II are satisfactory, though experiments with large scale test sets are still to be conducted, the results we get up to now are inspiring and totally support our argumentation.

When estimating samples in group-I, the accuracy is a little higher, which supports our first argumentation — it is a person's 3D facial structure that determines the appearance (intensity distribution) of his photo under variant illuminations, and a representative face (RF) can represent the 3D facial structure of all 3D face models contributing in computing it perfectly.

Though test images in group-II have nothing to do with the generation of the RF, we achieve comparable results when estimating samples in group-II. Actually the accuracy of group-II is only a tiny little lower than group-I in large (sometimes even a tiny little higher), which supports our second argumentation — there are some main types of facial structure, and the clustering technique does provide our illumination estimation system a good generalization ability.

Acknowledgements. The authors thank the Beijing University of Technology for providing the BJUT-3D Face Database. Portions of the research in this paper use the BJUT-3D Face Database collected under the joint sponsor of National Natural Science Foundation of China, Beijing Natural Science Foundation Program, Beijing Science and Educational Committee Program.

References

1. Claude, C.C., Fabrice, B.: Facial Expression Recognition: A Brief Tutorial Overview. *OnLine Compendium of Computer Vision* (2003)
2. Jiang, D.L., Hu, Y.X., Yan, S.C., Zhang, L., Zhang, H.J., Gao, W.: Efficient 3D Reconstruction for Face Recognition. *Pattern Recognition* 38(6), 787–798 (2005)
3. Blanz, V., Vetter, T.: Face Recognition Based on Fitting A 3D Morphable Model. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 25(9), 1063–1074 (2003)
4. Wu, J.W., Trivedi, M.M.: A Two-stage Head Pose Estimation Framework and Evaluation. *Pattern Recognition* 41(3), 1138–1158 (2008)
5. The BJUT-3D Large-Scale Chinese Face Database
6. Garland, M., Heckbert, P.S.: Surface Simplification Using Quadric Error Metrics. In: *Proc. of the SIGGRAPH 1997*, pp. 209–216. ACM Press, New York (1997)
7. Jebara, T.: Generating the Average 3D Face (2000), <http://www1.cs.columbia.edu/~jebara/htmlpapers/UTHEISIS/node48.html>
8. Chang, C.C., Lin, C.J.: LIBSVM: A Library for Support Vector Machines (2001), <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
9. Wang, H.M., Ou, Z.Y.: Face Recognition Based on Features by PCA/ICA and Classification with SVM. *Journal of Computer-Aided Design & Computer Graphics* 15(4), 416–420 (2003)

Bilateral Two-Dimensional Locality Preserving Projections with Its Application to Face Recognition

Xiao-Guo Wang

Institute of Communications Engineering, PLA Univ. of Sci. & Tech. ,
Nanjing 210007, China
imagewxg@163.com

Abstract. In this paper, we propose a novel algorithm for face feature extraction, namely the bilateral two-dimensional locality preserving projections (B2DLPP), which directly extracts the proper features from image matrices based on locality preserving criterion. Experiments on ORL and PIE face database are performed to test and evaluate the proposed algorithm. The results demonstrate the effectiveness of proposed algorithm.

Keywords: Locality preserving projections, Feature extraction, Face recognition.

1 Introduction

Reducing the dimensionality of the features space, i.e., feature extraction, is a common technique in statistical pattern recognition, typically used to lower the size of statistical models and overcome estimation problems, often resulting in an improved classifier accuracy in this lower-dimensional space. The most widely method are Eigenface[1] and Fisherface[1,2,3].

Locality preserving projections (LPP) is a recently proposed method in image recognition for feature extraction and dimension reduction[4]. The objective of LPP is to preserve the local structure of the image space by explicitly considering the manifold structure, which is in fact to solve a generalized eigenvalue problem (1)

$$XLX^T\xi = \lambda XDX^T\xi. \quad (1)$$

While, in many practical face recognition tasks, there are not enough samples to make XDX^T nonsingular, this is called a small sample size problem. One possible solution to attack this problem is to utilize the principal component analysis (PCA) as a preprocessing step to reduce the dimensionality of the vector space, which is known as Laplacianface algorithm and has been applied successfully to face representation and recognition[4,5,6].

However, in the existing Laplacianface (PCA+LPP) algorithm, some small principal components are thrown away in the PCA step, so some potential and valuable discriminatory information is lost in this step and the recognition performance are imperfect. Also, the 2D image matrices must be previously transformed into 1D image vectors. The resulting image vectors usually lead to a high-dimensional image vector space, where it is difficult to calculate the bases to represent the original

images, and such a matrix-to-vector transform may cause the loss of some structural information residing in original 2D images.

To avoid the disadvantages in the Laplacianface, Hu et al. proposed an alternative way to handle the above problem by directly projecting the image matrix under a specific projection criterion, rather than using the stretched image vector, which is a straightforward manner based on locality preserving criterion and the image matrix projection. Since the projection is a left-multiplying unilateral operation, so we called this method as left-multiplying unilateral two-dimension locality preserving projections (LU2DLPP)[7].

Inspired by Hu et al. [7], in this paper, the right-multiplying unilateral two-dimension locality preserving projections (RU2DLPP) and bilateral two-dimensional locality preserving projections (B2DLPP) are proposed to extract face feature by directly projecting the image matrix.

Experimental results on the ORL and PIE [8]face database show that the B2DLPP algorithm outperforms the Laplacianface, LU2DLPP and RU2DLPP algorithms in terms of the recognition performance rate.

2 LU2DLPP

Define the project equation as

$$y_i = U^T x_i . \tag{2}$$

where $U \in R^{m \times l}$ ($l \leq m$), $x_i \in R^{m \times n}$ is face image matrix. Like that of the vector-based LPP[5], the objective function of LU2DLPP is defined as

$$\min \sum_{i,j} \|y_i - y_j\|^2 W_{ij} . \tag{3}$$

The matrix W is a similarity matrix, a possible way of defining W is as follows: $W_{ij} = \exp(-\|X_i - X_j\|^2 / t)$, if x_i is among k-nearest neighbors of x_j or x_j is among k-nearest neighbors of x_i , otherwise, $W_{ij} = 0$. Since $\|A\|^2 = tr(AA^T)$, so

$$\begin{aligned} \frac{1}{2} \sum_{ij} \|y_i - y_j\|^2 W_{ij} &= \frac{1}{2} \sum_{ij} tr((y_i - y_j)(y_i - y_j)^T) W_{ij} \\ &= \frac{1}{2} \sum_{ij} tr(y_i y_i^T + y_j y_j^T - y_i y_j^T - y_j y_i^T) W_{ij} \\ &= tr\left(\sum_i D_i y_i y_i^T - \sum_{ij} W_{ij} y_i y_j^T\right) \\ &= tr\left(\sum_i D_i U^T x_i x_i^T U - \sum_{ij} W_{ij} U^T x_i x_j^T U\right) \end{aligned}$$

$$\begin{aligned}
&= \text{tr} \left(\sum_i D_{ii} U^T x_i x_i^T U - \sum_{ij} W_{ij} U^T x_i x_j^T U \right) \\
&= \text{tr} \left(U^T \left(\sum_i D_{ii} x_i x_i^T - \sum_{ij} W_{ij} x_i x_j^T \right) U \right) \\
&= \text{tr} \left(U^T \left(\sum_i D_{ii} x_i x_i^T - \sum_{ij} W_{ij} x_i x_j^T \right) U \right) \\
&= \text{tr} (U^T X (D - W) X^T U). \tag{4}
\end{aligned}$$

Therefore, a constraint $U^T XDX^T U = 1$ is added, the minimization problem becomes generalized eigenvalue problem

$$XLX^T \xi = \lambda XDX^T \xi. \tag{5}$$

Where $L = D - W$, $X = [x_1, x_2, \dots, x_N]$. Let the column vectors u_1, u_2, \dots, u_l be the solutions of Eq. (5), ordered according to their eigenvalues $\lambda_1 < \lambda_2, \dots, \lambda_l$. Thus the projection matrix is as follows: $U = [u_1, u_2, \dots, u_l]$. According to Eq. (2), the face image feature y_i is a $l \times n$ matrix ($l < m$).

3 RU2DLPP

Define the project equation as

$$y_i = x_i V. \tag{6}$$

Where $V \in R^{n \times d}$ ($d \leq n$). Similar to LU2DLPP, the Eq. (3) minimization problem becomes generalized eigenvalue problem

$$X^T L X \xi = \lambda X^T D X \xi. \tag{7}$$

Let the column vectors v_1, v_2, \dots, v_d be the solutions of Eq. (7), ordered according to their eigenvalues $\lambda_1 < \lambda_2, \dots, \lambda_d$. Thus the projection matrix is as follows: $V = [v_1, v_2, \dots, v_d]$. According to Eq. (6), the face image feature y_i is a $m \times d$ matrix. We called this method as the right-multiplying unilateral two-dimension locality preserving projections (RU2DLPP).

4 B2DLPP

As described in section two and three, we can see that the LU2DLPP and RU2DLPP just extract features according to column or row of image matrix individually. For the

purpose of extract features according to column and row of image matrix simultaneity, we proposed the B2DLPP algorithm.

Define the project equation as

$$y_i = U^T x_i V \tag{8}$$

Where $U \in R^{m \times l}$ ($l \leq m$), $V \in R^{n \times d}$ ($d \leq n$), since $\|A\|^2 = tr(AA^T)$, so

$$\begin{aligned} \frac{1}{2} \sum_{ij} \|y_i - y_j\|^2 W_{ij} &= \frac{1}{2} \sum_{ij} tr\left((y_i - y_j)(y_i - y_j)^T\right) W_{ij} \\ &= tr\left(\sum_i D_{ii} y_i y_i^T - \sum_{ij} W_{ij} y_i y_j^T\right) \\ &= tr\left(\sum_i D_{ii} U^T x_i V V^T y_i^T - \sum_{ij} W_{ij} U^T x_i V V^T x_j^T U\right) \\ &= tr\left(U^T \left(\sum_i D_{ii} x_i V V^T y_i^T - \sum_{ij} W_{ij} U^T x_i V V^T x_j^T\right) U\right) \\ &= tr\left(U^T (D_v - W_v) U\right) \end{aligned} \tag{9}$$

Where $D_v = \sum_i D_{ii} x_i V V^T x_i^T$, $W_v = \sum_{ij} W_{ij} x_i V V^T x_j^T$. The constraint equation is $U^T x_i V \sum_i D_{ii} V^T x_i^T U = U^T D_v U = I$, the Eq. (9) becomes generalized eigenvalue problem:

$$(D_v - W_v)u = \lambda D_v u \tag{10}$$

Similarly, we can get generalized eigenvalue problem

$$(D_u - W_u)v = \lambda D_u v \tag{11}$$

Now we discuss how to solve the optimization problems (10) and (11). It is easy to see that the optimal U should be the generalized eigenvectors of $(D_v - W_v, D_v)$ and the optimal V should be the generalized eigenvectors of $(D_u - W_u, D_u)$. Since the matrices D_v , W_v , D_u , W_u are not fixed, we can not compute the optimal U and V simultaneously. In this paper, we compute U and V iteratively as follows. Firstly, we fix V , then U can be computed by solving the Eq. (10). Once U is obtained, V can be updated by solving the Eq. (11). Thus, the optimal U and V can be obtained by iteratively computing the generalized eigenvectors of (10) and (11). In our experiments, U is initially set to the identity matrix.

5 Experiments

To demonstrate the efficiency of our method, extensive experiments are done on the ORL and PIE database. All feature extraction methods are compared on the same training sets and testing sets. The classifier is nearest neighbor, the weight matrix W is defined by cosine: $W_{ij} = \cos(x_i, x_j)$ if x_i and x_j are in the same class, otherwise $W_{ij} = 0$.

For the ORL database, there are 40 individuals, each individual has 10 samples, and the sample size is 112×92 , we randomly choose 200 samples (5 for each individual) as the training set, the remaining 200 samples are used as the test set. For the PIE database, we construct a sub-database based on PIE. In the sub-database, there are 68 individuals, each individual has 20 samples, and the sample size is 32×32 , we randomly choose 680 samples (10 for each individual) as the training set, the remaining 680 samples are used as the test set. Such procedures are repeated for 30 times, which results in 30 groups of data. The results shown in the following tables are the averages of 30 times (the “i” is the iterative times).

The recognition results on the ORL database are shown in Table 1. It is found that the 2DLPP (LU2DLPP, RU2DLPP and B2DLPP) methods significantly outperform the 1DLPP (Laplacianface) method, and the B2DLPP performance is the best with the iterative times equal to 2, which gets the lowest recognition error rate at 2.8% with number 6×9 of features.

The recognition results on the PIE sub-database are shown in Table 2. It is found that the 1DLPP performance is superior to LU2DLPP but worse than RU2DLPP, while B2DLPP outperforms 1DLPP, LU2DLPP and RU2DLPP methods. The B2DLPP method gets the best performance with a recognition error rate of 21.8%, with number 20×4 of features.

According to Table 1 and Table 2, it is found that the optimal U and V can be obtained by 2 or 3 iterative times.

Table 1. Performance comparison on ORL database

Feature extraction method	1DLPP	LU2DLPP	RU2DLPP	B2DLPP (i=1)	B2DLPP (i=2)
Feature dimension	40	10×92	112×8	4×7	6×9
Recognition error rate (%)	7.5	3.9	4.1	3.0	2.8
Feature extraction method	B2DLPP (i=3)	B2DLPP (i=4)	B2DLPP (i=5)	B2DLPP (i=6)	B2DLPP (i=10)
Feature dimension	6×8	6×8	6×9	6×8	6×8
Recognition error rate (%)	2.9	2.9	2.9	2.9	2.9

Table 2. Performance comparison on PIE sub-database

Feature extraction method	1DLPP	LU2DLPP	RU2DLPP	B2DLPP (R=1)	B2DLPP (R=2)
Feature dimension	143	11×32	32×12	14×5	17×6
Recognition error rate (%)	28.7	26.7	33.6	23.8	22.3
Feature extraction method	B2DLPP(R=3)	B2DLPP(R=4)	B2DLPP(R=5)	B2DLPP(R=6)	B2DLPP(R=10)
Feature dimension	20×4	18×5	17×5	16×5	16×5
Recognition error rate (%)	21.8	21.9	22.0	22.1	22.0

6 Conclusion

In this paper, we present a novel LPP-based subspace projection method for dimensionality reduction and feature extraction that we refer to as B2DLPP. The B2DLPP works directly on the image matrix of images and extract the face feature bilateral. Experimental results on the ORL and PIE face database show the effectiveness of the proposed method. It is believed that the proposed algorithm here should stimulate a much wider use of LPP-based approaches.

References

1. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Trans. Pattern Analysis and Machine Intelligence* 19(7), 711–720 (1997)
2. Zhao, W., Chellappa, R., Phillips, P.J.: Subspace Linear Discriminant Analysis for Face Recognition. Tech Report CAR-TR-914, Center for Automation Research, University of Maryland (1999)
3. Swets, D.L., Weng, J.: Using Discriminant Eigenfeatures for Image Retrieval. *IEEE Trans. Pattern Analysis and Machine Intelligence* 18(8), 831–836 (1996)
4. Belkin, M., Niyogi, P.: Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Computation* 15(6), 1373–1396 (2003)
5. He, X.F., Niyogi, P.: Locality Preserving Projections. In: *Advances in Neural Information Processing Systems*, vol. 16. MIT Press, Cambridge (2003)
6. He, X.F., Yan, S.C., Hu, Y.X., Niyogi, P., Zhang, H.J.: Face Recognition Using Laplacian-faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(3), 328–340 (2005)
7. Hu, D.W., Feng, G.Y., Zhou, Z.T.: Two-dimensional Locality Preserving Projections (2DLPP) with Its Application to Palmprint Recognition. *Pattern Recognition* 40(1), 339–342 (2007)
8. Sim, T., Baker, S., Bsat, M.: The CMU Pose, Illumination, and Expression (PIE) Database. In: *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition* (May 2002)

DT-CWT Feature Structure Representation for Face Recognition under Varying Illumination Using EMD

Yuehui Sun¹ and Di Zhang²

¹ School of Information and Engineering, Guangdong University of Technology, Guangzhou 510006, China

² College of Computer Science, Shaoguan University, Shaoguan 512005, China
yuehui_sun@126.com

Abstract. We introduce a method for illumination detection and removal technique using Empirical Mode Decomposition (EMD) to decompose subimages of Dual-Tree Complex Wavelet Transform (DT-CWT). The subimages are reconstructed without illumination distortion components for face recognition. Compared with others, this method has the following advantages: it can be directly applied without any prior information; it has perfectly reconstruction ability because of DT-CWT in low frequency. Experiments are carried out upon the Yale B and CMU PIE face databases, and the results demonstrate that the proposed method shows satisfactory recognition rates under varying illumination conditions.

Keywords: Face recognition, EMD, DT-CWT.

1 Introduction

Variable illumination is one of the most important problems in face recognition. Although, the ability of algorithms to recognize faces across illumination changes has made strong progress in the last years. Evaluations such as the Face Recognition Vendor Test (FRVT) [1] and the Face Recognition Grand Challenge (FRGC) [2] indicate that illumination still has an important effect on the recognition process. It has been shown experimentally and theoretically that illumination variations can cause a significant degradation in performance of facial recognition systems [3]. In the past few years, many methods have been proposed to solve this problem with improvements in recognition [4-7]. However, these kinds of approaches have two main drawbacks. First, the different representations of image can be only extracted once and the new images of a different person which is not included in the training set cannot be handled. Second, features for identity are weakened when the illumination-invariant features are extracted. For example, some authors suggest that illumination normalization can be achieved by discarding the first three PCA components. The complexity and assumptions of idealities in many of these methods often limit its application in practical problems.

In this paper we will show the power of the combination EMD and DT-CWT in addressing illumination removal effects in face recognition. Firstly, we get subimages

which encompass information of different spatial frequency, spatial localities and orientations by DT-CWT. Using EMD to decompose two dimensional DT-CWT facial subimages into their fundamental source signals, we can isolate the effects of illumination to one or more of these source signals. By reconstructing the sub-image without these illumination artifact source signals, we can reduce the overall effect of illumination variation. Both implementations of DT-CWT and EMD are simple while still being effective. Recognition results to demonstrate the improvement of DT-CWT combined EMD processing are reported using PCA, LDA, and especially improved ONPP on the Carnegie Mellon University Pose-Illumination-Expression (CMU PIE) database [8].

2 DT-CWT and EMD Algorithm Review

The Dual-Tree Complex Wavelet Transform (DT-CWT) [9] as a wavelet transform proposed and studied which has been found to be particularly suitable for image decomposition and representation when the goal is the derivation of local and discriminating features like Gabor wavelet [10]. DT-CWT provides good directional selectivity in six different fixed orientations at some scales, which is able to distinguish positive and negative frequencies. And it has a limited redundancy of four for images and is much faster than the Gabor wavelet to compute. Therefore, DT-CWT filter representation gives better performance for classifying facial actions. The DT-CWT expansion of an image $f(\vec{x})$ is given by:

$$f(\vec{x}) = \sum_k W_\phi(j_o, k) \phi_{j_o, k}(\vec{x}) + \sum_i \sum_{j > j_o} \sum_k W_\psi(j, k) \psi_{j, k}^i(\vec{x}) \quad (1)$$

where $i = \pm 15^\circ, \pm 45^\circ, \pm 75^\circ$. The scaling function $\phi_{j_o, k}$ and wavelet function $\psi_{j, k}^i$ are complex. $W_\phi(j_o, k)$ indicates the scaling coefficients and $W_\psi(j, k)$ are the wavelet coefficients of the transform. Hence six sub-bands are obtained corresponding to the direction $i = \pm 15^\circ, \pm 45^\circ, \pm 75^\circ$, as showed in Fig. 1.

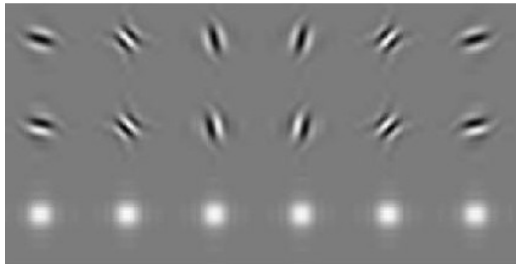


Fig. 1. Filter impulse response of DT-CWT in frequency domain

A general overview of EMD and its implementation is presented in [11], but we will briefly summarize EMD here. The aim of the EMD is to decompose the signal into a sum of intrinsic mode functions (IMFs). The following algorithm defines this procedure and outlines most EMD implementations. Given a source signal $x(t)$:

1. Found all local extrema of $x(t)$
2. Interpolate between all minima(maxima) to get a envelope $e_{\min}(t)$ ($e_{\max}(t)$) and
 compute the mean envelope $m(t) = (e_{\min}(t) + e_{\max}(t)) / 2$
3. Extract detail $d(t) = x(t) - m(t)$
4. Check if $|m(t)| < \epsilon$, if not repeat step 1-3 with $d(t)$ instead of $x(t)$; if so, $d(t)$
 is an IMF
5. Calculate residual $r(t) = x(t) - d(t)$
6. Return to step 1 with $r(t)$ as $x(t)$
7. Repeat until signal has no extreme

By simply summing all the IMF's together we will recover the original data accommodating for minor variations due to the interpolation present in the algorithm. EMD also allows us to selectively reconstruct the data, ignoring the IMF's whose contributions to the data are undesirable. For our application, such contributions are those of illumination effects. If we can use EMD to decompose our original facial images into their IMF's, there is a strong likelihood that the effects of illumination will be isolated to one or more IMF's. Selective reconstruction of facial images using IMF's that do not contain illumination effects will enable us to reconstruct the data without the unwanted effects of illumination variation.

3 Improved ONPP

ONPP [12] is a new linear dimensionality reduction algorithm. Compared to other dimensionality reduction techniques, ONPP can be viewed as a synthesis of PCA and LLE [13]. ONPP is a linear method, while Isomap [14] and LLE are nonlinear methods, so neither of them can deal with new test data points except training data points. The main idea of ONPP is to seek an orthogonal mapping of a given data set $X = (x_1, x_2, \dots, x_n) \in R^{m \times n}$ so as to best preserve a graph which describes the local geometry.

The reconstruction errors are measured by minimizing the objective function:

$$\mathcal{E}(W) = \sum_i \left\| y_i - \sum_j w_{ij} y_j \right\|_2^2 \quad (2)$$

In the undersampled size case where m is greater than n , Kokiopoulou and Saad prove that the rank of M is $n-c$ (c is the number of classes). In order to ensure that the resulting matrix M will be nonsingular, ONPP employs an initial PCA projection that

reduces the dimensionality of the data vectors to $n-c$ like LDA. In the following we introduce improved ONPP (IONPP) to overcome the problem.

Given a set of training images A_i and we construct a graph by k-NN, using the Frobenius norm for measure. It is easy to prove that $\|A_i\|_F^2 = \|x_i\|_2^2$, where x_i is the column vector of A_i . The neighbor graph can be constructed as the same as ONPP, also the weights matrix W can be figured out the same way as IONPP.

Considering projecting an $m \times n$ image A_i into the m -dimensional Euclidean space, we get the equation:

$$y_i = A_i v \tag{3}$$

Similar to the objective function in ONPP, we have

$$\min \sum_i \left\| y_i - \sum_j w_{ij} y_j \right\|_2^2 \tag{4}$$

That means if A_i and A_j are close in the high dimension spaces, their projections will keep the affinity in the reduced spaces. The objective function can be written as

$$\begin{aligned} & \sum_i \left\| y_i - \sum_j w_{ij} y_j \right\|_2^2 \\ &= \sum_i \left\| A_i v - \sum_j w_{ij} A_j v \right\|_2^2 \\ &= \sum_i \left\| (A_i - \sum_j w_{ij} A_j) v \right\|_2^2 \\ &= \sum_i v^T (A_i^T - \sum_j w_{ij} A_j^T) (A_i - \sum_j w_{ij} A_j) v \\ &= \sum_i v^T (A_i^T A_i - \sum_j A_j^T w_{ij} A_i - \sum_j A_i^T w_{ij} A_j + \sum_{m,n} A_m^T w_{im} w_{in} A_n) v \\ &= v^T (\sum_i A_i^T A_i - \sum_{i,j} A_j^T w_{ij} A_i - \sum_{i,j} A_i^T w_{ij} A_j + \sum_{k,i,j} A_i^T w_{ki} w_{kj} A_j) v \\ &= v^T A^T [(I - W^T)(I - W) \otimes I_m] A v \end{aligned} \tag{5}$$

where $A = [A_1^T, A_2^T, \dots, A_n^T]^T$, let $U = (I - W^T)(I - W)$, furthermore, to remove an arbitrary scaling factor in the embedding, we impose a constraint on v :

$$v^T A^T (I \otimes I_m) A v = 1 \tag{6}$$

So the objective function becomes

$$\arg \min_{v^T A^T (I \otimes I_m) A v = 1} v^T A^T (U \otimes I_m) A v \quad (7)$$

The solution of v is given by the minimum eigenvectors of the generalized eigenvalue problem as below

$$A^T (U \otimes I_m) A v = \lambda A^T (I \otimes I_m) A v \quad (8)$$

The algorithm is summarized in the following:

1. Constructing the affinity graph Let $G(A', E)$ denote a graph with n nodes $A' = (A_1, A_2, \dots, A_n)$, E is the edges between each nodes. An edge $e_{ij} = (A_i, A_j)$ exists if A_i is one of the k nearest neighbors of A_j , or A_i and A_j are of the same class (supervised IONPP).
2. Compute the weight w_{ij} using methods in ONPP which gives the best linear construction of each data point A_i by its neighbors.
3. Compute matrix V whose column vectors are the d eigenvectors of

$$A^T (U \otimes I_m) A v = \lambda A^T (I \otimes I_m) A v$$

corresponding to the first d smallest eigenvalues.

4. Compute the projections of data points by $Y_i = A_i V$

4 Preprocessing Combined DT-CWT with EMD

As Fig. 2 showed, the magnitude response of DT-CWT for a face image of Fig. 4(a) is given. Considering the size of the face image showed in Figure 4(a), is 128×128 , using DT-CWT with 4 levels and 6 directions, this provides twenty-four sub-bands which encompasses information of different spatial frequency, spatial localities and orientations of face feature.

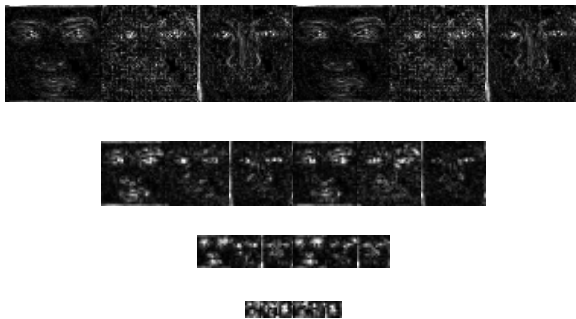


Fig. 2. The magnitude of response using DT-CWT for levels $N=1, 2, 3, 4$ of the orientation

Illumination effects are primarily due to one primary source of light which creates the majority of shadows such as in Fig. 4(a) and (c). As such we can treat these effects as linear in one dimensional sense, although not necessarily along the typical axes. Treating each row or column of a facial image as separable we can string two-dimensional facial images into one-dimensional vectors. Application of EMD to vectors transformed by DT-CWT yields a set of vector IMF's which are then reshaped into matrix IMF's as showed in Fig. 3.

The stopping conditions set in the EMD algorithm determine the exact number of IMF's but for our experiments and data we found that we obtain thirteen IMF's. Regardless of the exact number of IMF's, the last two IMF's contains the majority of the illumination effects. Due to nature of the EMD algorithm, as the order of the IMF increases the relative mean of the data approaches zero [15]. As such, by applying EMD to facial images that are subject to illumination effects we can partition the effects into two types, shadowing and reflections. Since shadowing darkens regions of an image, it creates low-valued regions while reflections create relatively high-valued regions. These are effectively the largest magnitude extreme in the images, but also most slowly changing. In other words they represent the lower spatial frequency contents of the image. EMD isolates these frequencies in the last few IMF's.

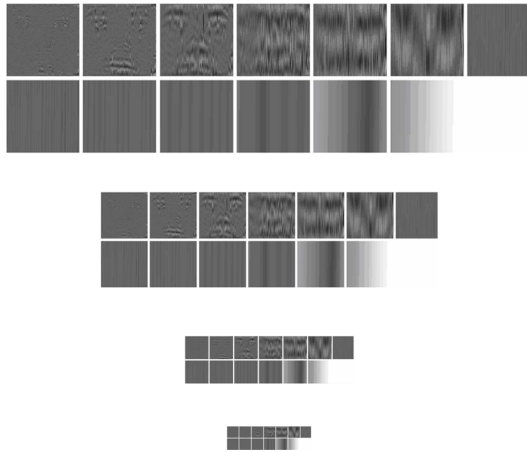


Fig. 3. Resulting DT-CWT IMF's from Figure 4 (a). Ordered from left to right, top to bottom in increasing order, each level using same orientation.

With this in mind, we look at the last two IMF's and determine which one introduced the shadowing artifacts to the data. This is easily done by comparing the means of the two IMF's and choosing the smaller one. Once we have determined which IMF is responsible for the effects of shadowing, we reconstruct the image without that IMF. The resulting facial image showed in Fig. 4(b) and (d) now contains significantly less shadowing effects and allows the fundamental nature of the facial image to come through more.

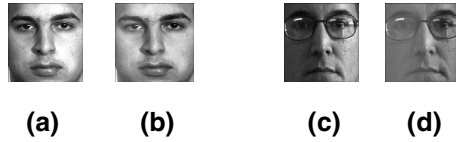


Fig. 4. Examples of facial image reconstruction excluding unwanted IMF. (a) Original image with right-side reflection. (b) Reconstructed image minus reflection. (c) Original image with left-side cast shadow. (d) Reconstructed image minus left-side cast shadow effects.

5 Recognition Results

The PIE (Pose, Illumination, and Expression) database contains a total of 41,368 images from 68 individuals, with different pose, illumination and expression conditions. The images were taken using the CMU 3D room using a set of 13 synchronized high-quality color cameras and 21 flashes. The images are classified into different sets depending on the including pose, illumination and expression variations. Regarding illumination conditions, PIE takes into account the fact that in the real world, illumination usually consists of an ambient light with perhaps one or two point sources. To obtain representative images of such cases PIE creators decided to capture images with the room lights on and with them off. For our experiments we use only frontal images, which corresponds to the ones captured using camera c27 (in the PIE terminology). Considering DT-CWT convenience, we change image size to 128×128.

Table 1. Average Equal Error Rate (EER) for PCA and LDA using EMD combined with DT-CWT processed images

Num. of Training Images	PCA	DT-CWT +EMD+PCA	LDA	DT-CWT +EMD+LDA
2	0.2462	0.0534	0.2210	0.2569
3	0.1567	0.0380	0.1724	0.2235
4	0.1590	0.0300	0.1670	0.1654
5	0.1314	0.0213	0.1245	0.1325
6	0.1875	0.0176	0.1625	0.1530
7	0.1345	0.0160	0.1344	0.1235
8	0.1510	0.0120	0.1023	0.0997
9	0.1320	0.0084	0.1260 9	0.0958
10	0.178	0.0097	0.1210	0.0968
11	0.2012	0.0095	0.1320	0.1315

Applying our EMD preprocessing to the entire database removed the significant illumination variation from the facial images. Training sets vary in size and composition by random selection over multiple experiments for each of the four recognition algorithms using Euclidean distance measure and nearest neighbor classifier. Each experiment involved training the recognition algorithm using the specified training set and then recording verification results. For PCA [16], LDA, ONPP and IONPP, ten experiments were run each. Performance is quantified by average Equal Error Rate (EER) over all experiments.

Experimental results showed in Tab. 1 demonstrate that EMD reprocessing is an effective approach in normalizing a facial image in both space and frequency especially when using very small sample size training sets. In all of the cases, the pre-processing algorithm increases the recognition rate. Second conclusion is that the highest recognition rates are obtained by PCA while LDA not. We do not have a clear explanation for this phenomenon, but one of the possible reasons is the different similarity metrics (e.g. Hamming distances) should be used in these cases.

For each experiment we used a fixed number of training images per individual, 2 to 10. In order to obtain representative results we take the average of several sets of experiments for each fixed number of training images.

In order to show advantages of IONPP easily, we take two images of each subject are randomly chosen for training, while the remaining one is used for testing. One can see from Fig. 5 that IONPP performs the better than ONPP, no matter using preprocessing or not. Both ONPP and IONPP are good at representing data, but are restrained for great variations in lighting without any preprocessing. However, both the face recognition algorithms using EMD combined with DT-CWT to remove illumination improved the performance significantly.

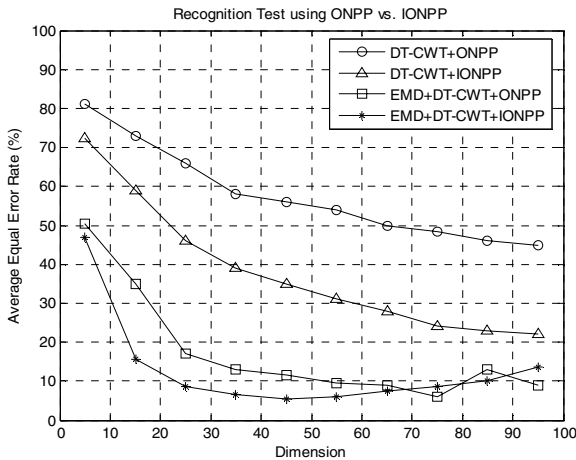


Fig. 5. Face recognition experiments using ONPP and IONPP with illumination removal preprocessing using 2 images randomly selected in training set

Fig. 6 shows the average recognition rate of the four different approaches in a bar chart, where the center tick of each bar denotes the average error rate and the intervals correspond to the minimum and maximum values of error rates. For a fair comparison, we used the number of training images and dimension reduced of each approach such that it produces the best recognition rate. This figure shows that the method proposed not only outperforms the other methods but also has the smallest variance of recognition rate over the 10 runs. This implies that our method performs very robust and stable face recognition irrespective of the change of lighting conditions.

In summary, by analyzing the simulations carried out using PIE databases we conclude that some of the compared algorithms achieve very high recognition rates when used as a pre-processing stage of standard eigenspace-based face recognition systems.

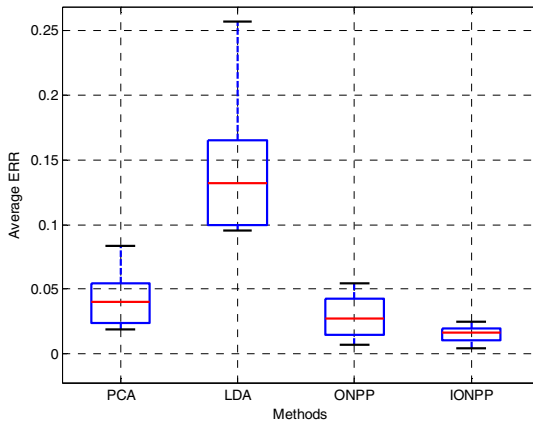


Fig. 6. Recognition results of four different methods using illumination removal method: PCA, LDA, ONPP and IONPP

6 Conclusions

Variable illumination is a major problem in face recognition. We concentrated ourselves in algorithms with the following properties: general purpose, no modeling steps or training images required, simplicity, high speed, and high performance in terms of recognition rates. Experimental results demonstrate that EMD combined with DT-CWT preprocessing does not introduce undesired noise to the images with no a priori information. Furthermore, IONPP shows better recognition effects than original ONPP. The simple implementation and effectiveness of EMD preprocessing indicates its usefulness as a preprocessing step in facial recognition algorithm. Expanding on the work presented here, we plan to improve results through the use of a true two-dimensional EMD algorithm that can possibly be more suited when dealing from illumination artifacts arising from more than one illumination source.

Acknowledgements. The research described in this paper was supported by NSFC (No. U60772117), the Scientific Research Starting Foundation for Doctorate, Guangdong University of Technology, China (Grant No. 083033).

References

1. Phillips, P.J., Grother, P., Micheals, R., Blackburn, D.M., Tabassi, E., Bone, M.: Face Recognition Vendor Test 2002, Evaluation report. Technical Report IR 6965, NIST (2003)
2. Phillips, P.J., Flynn, P.J., Scruggs, T., Bowyer, K.W., Chang, J., Hoffman, K., Marques, J., Jaesik, M., Worek, W.: Overview of the Face Recognition Grand Challenge. In: CVPR 1, pp. 947–954 (2005)
3. Adini, Y., Moses, Y., Ullman, S.: Face Recognition: The Problem of Compensating for Changes in Illumination Direction. *IEEE Trans. Pattern Analysis and Machine Intelligence* 19, 721–732 (1997)
4. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Trans. Pattern Analysis and Machine Intelligence* 19, 711–720 (1997)
5. Chen, T., Hsu, Y.J., Liu, X., Zhang, W.: Principle Component Analysis and Its Variants for Biometrics. In: *Image Processing Proceedings International Conference*, vol. 1, pp. 61–64 (2002)
6. Gross, R., Matthews, I., Baker, S.: Eigen Light-Fields and Face Recognition Across Pose. In: *5th IEEE International Conference Automatic Face and Gesture Recognition Proceedings*, pp. 1–7 (2002)
7. He, X., Niyogi, P.: Locality Preserving Projections. In: *Advances in Neural Information Processing Systems*, vol. 16, pp. 153–160 (2003)
8. Sim, T., Baker, S., Bsat, M.: The CMU Pose, Illumination, and Expression Database. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25, 1615–1618 (2003)
9. Kingsbury, N.G.: The Dual-tree Complex Wavelet Transform: a New Efficient Tool for Image Restoration and Enhancement. In: *Proc. European Signal Processing Conf.*, pp. 319–322 (1998)
10. Liu, C.J., Wechsler, H.: Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition. *IEEE Trans. Image Processing* 11, 467–476 (2002)
11. Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C., Liu, H.H.: The Empirical Mode Decomposition and Hilbert Spectrum for Nonlinear and Nonstationary Time Series Analysis. *Proc. Royal Society of London A* 454, 903–995 (1998)
12. Kokiopoulou, E., Saad, Y.: Orthogonal Neighborhood Preserving Projections. In: *IEEE Int. Conf. on Data Mining*, pp. 1–8 (2005)
13. Roweis, S., Saul, L.: Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science* 290, 2323–2326 (2000)
14. Tenenbaum, J.B., de Silva, V., Langford, J.C.: A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science* 290, 2319–2323 (2000)
15. Flandrin, P., Goncalves, P., Rilling, G.: Empirical Mode Decomposition as a Filter Bank. *IEEE Signal Processing Letters* 11, 112–114 (2004)
16. Turk, M.A., Pentland, A.P.: Face Recognition Using Eigenfaces. In: *Computer Vision and Pattern Recognition Proceedings IEEE Computer Society Conference*, pp. 586–591 (1991)

Spatially Smooth Subspace Face Recognition Using LOG and DOG Penalties

Wangmeng Zuo¹, Lei Liu¹, Kuanquan Wang¹, and David Zhang²

¹ School of Computer Science and Technology, Harbin Institute of Technology
150001 Harbin, China

{wmzuo, wangkq}@hit.edu.cn

² Department of Computing, the Hong Kong Polytechnic University
Kowloon, Hong Kong

csdzhang@comp.polyu.edu.hk

Abstract. Subspace face recognition methods have been widely investigated in the last few decades. Since the pixels of an image are spatially correlated and facial images are generally considered to be spatially smoothing, several spatially smooth subspace methods have been proposed for face recognition. In this paper, we first survey the progress and problems in current spatially smooth subspace face recognition methods. Using the penalized subspace learning framework, we then proposed two novel penalty functions, Laplacian of Gaussian (LOG) and Derivative of Gaussian (DOG), for subspace face recognition. LOG and DOG penalties introduce a scale parameter, and thus are more flexible in controlling the degree of smoothness. Experimental results indicate that the proposed methods are effective for face recognition, and achieve higher recognition accuracy than the original subspace methods.

Keywords: Subspace analysis, Face recognition, Regularization, Laplacian of Gaussian, Derivative of Gaussian.

1 Introduction

Face recognition have been one of the most important issues in computer vision and pattern recognition over the last several decades [1]. Since of its simplicity and good generalization, subspace analysis approaches has received considerable research interests and has been widely investigated in face recognition. Currently, varieties of subspace methods, such as Eigenfaces, Fisherfaces, Laplacianfaces, independent component analysis, and manifold learning, have been proposed and applied to face recognition tasks [2, 3, 4].

Most subspace face recognition approaches should derive the projection vectors by solving the generalized eigenvalue problem of two (scatter) matrices \mathbf{W} and \mathbf{D} . However, when applied to face recognition, since the number of the available training samples is limited, the projection vectors usually would be overfitted to the training set. To improve the generalization performance of subspace methods, by far, a number of modification approaches have been recently proposed, which can be roughly grouped into four categories, regularized subspace analysis, tensor subspace analysis,

post-processed subspace analysis, and penalized subspace analysis. In the early regularized subspace analysis approaches, researchers did not notice the importance of preserving the spatial smoothness of the projection vectors, and only used several standard regularization techniques to alleviate the poor estimation of the matrix \mathbf{W} and \mathbf{D} [5, 6, 7].

In tensor subspace analysis, each facial image is regarded as a two-order tensor, and multilinear singular value decomposition techniques could then be used to compute the row and column projection matrices [8, 9, 10, 11]. Actually, given N training images with size $m \times n$, the number of the samples would be much higher than N , and the size of the scatter matrices would be much lower than $mn \times mn$ during the calculation of the row and column projection matrices. Thus, tensor subspace analysis is expected to have a good generalization performance for face recognition. However, in the calculation of the column projection matrix, tensor subspace analysis usually neglects the spatial relation between rows. Analogously, in the calculation of the row projection matrix, the spatial relation between columns is neglected.

Recently, a kind of post-processing approach has been proposed to directly smooth the projection vectors using a circular Gaussian filter [12]. In [13], Hao et al. further proved the equivalence of the post-processing approach and the image Euclidean distance (IMED) method [14]. However, there are not any constraints in the stage of spatially smoothing, and by far only the Gaussian filters are empirically chosen in the post-processing approach.

Penalized subspace analysis can also be adopted for spatially smoothing subspace learning. In [15], Hastie et al. proposed a penalized discriminant analysis method using the Laplacian penalty. Most recently, Cai et al. introduced a generalized penalized subspace learning model for almost all existing subspace methods. Penalized subspace analysis uses the same procedure as regularized subspace analysis to improve the generalization performance. As to the regularization term, penalized subspace analysis [16] adopted the Laplacian penalty while regularized subspace analysis usually adopted the identity matrix or the diagonal matrix. Since the Laplacian penalty improves the poor matrix estimation by taking into account the spatially smoothing prior of images, penalized subspace analysis is expected to be effective in achieving better generalization performance.

The Laplacian penalty function in Cai's method, however, does not have any scale parameters, and is inflexible in changing the degree of smoothness for facial images with different resolution. In this paper, we proposed two novel penalty functions, LOG and DOG, for subspace face recognition. The two penalty functions introduce a scale parameter to treat the degree of smoothness. Our experimental results indicate that the proposed methods are effective for face recognition, and achieve higher recognition accuracy than the original subspace face recognition methods.

The remainder of the paper is organized as follows: Section 2 presents a survey on four spatially smooth subspace face recognition approaches, regularized subspace analysis, tensor subspace analysis, post-processed subspace analysis, and penalized subspace analysis. Section 3 proposes two novel penalty functions, LOG and DOG, for subspace face recognition. In Section 4, experiments are used to evaluate the proposed method. Finally, Section 5 concludes this paper.

2 Spatially Smooth Subspace Face Recognition

In this section, we first introduce the procedure of the subspace methods, and then present a survey on four kinds of spatially smooth subspace face recognition approaches, regularized subspace analysis, tensor subspace analysis, post-processed subspace analysis, and penalized subspace analysis.

Let $\{(\mathbf{X}_i, y_i) | i=1, 2, \dots, N\}$ denote a training set, where \mathbf{X}_i is the i th facial image with image size $m \times n$, and y_i is the corresponding class label of \mathbf{X}_i . In some subspace analysis methods, the image \mathbf{X}_i must be concatenated into a 1D vector \mathbf{x}_i in advance. Thus, we also use $\{(\mathbf{x}_i, y_i) | i=1, 2, \dots, N\}$ to denote the same training set, where $\mathbf{x}_i \in \mathbb{R}^{mn}$, and let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$ to denote the set of the training data. We assume that the mean of the training samples is zero.

A number of subspace analysis approaches, such as Fisher's linear discriminant (LDA), discriminative common vectors (DCV), locality preserving projection (LPP), and margin Fisher analysis (MFA), have been proposed and applied to face recognition. Actually, all these methods can be generally represented in one graph embedding framework defined as

$$\mathbf{w}^* = \arg \max_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{X} \mathbf{W} \mathbf{X}^T \mathbf{w}}{\mathbf{w}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{w}}, \quad (1)$$

where \mathbf{W} would be defined as the weighted matrix and \mathbf{D} would be defined as the diagonal matrix [16, 17]. With different definitions of \mathbf{W} and \mathbf{D} , the general framework would lead to different subspace analysis methods. For example, LDA can be represented using this framework if we define

$$W^{LDA}(i, j) = \begin{cases} 1/m_i, & \text{if } \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ belong to the } t\text{th class} \\ 0, & \text{otherwise} \end{cases}, \quad (2)$$

$$D^{LDA}(i, j) = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}, \quad (3)$$

where N_t is the number of the training samples which belong to the t th class. In the following, we will present a survey of these four kinds of spatially smooth subspace analysis approaches using the general graph embedding framework.

2.1 Regularized Subspace Analysis

Following the general graph embedding framework, regularized subspace analysis added a small perturbation on the diagonal matrix to improve the generalization performance. Let

$$\Xi_D = \mathbf{X} \mathbf{D} \mathbf{X}^T = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T, \quad (4)$$

where $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_d]$ are the d eigenvectors of the matrix Ξ_D , and $\mathbf{\Lambda}$ is a diagonal matrix with the corresponding eigenvalues $\lambda_i = \mathbf{\Lambda}(i, i)$. Usually, regularized subspace analysis [5, 6] uses the regularized matrix

$$\Xi_D^R = \mathbf{U}\Lambda_R\mathbf{U}^T, \tag{5}$$

to replace the matrix defined in Eq. (4), where

$$\Lambda_R = \Lambda + \lambda\mathbf{I}. \tag{6}$$

More complicatedly, Dai et al. [7] proposed a three-parameter regularization method on the matrix Ξ_D by defining the regularized diagonal matrix $\Lambda_R^{\alpha,\beta,\gamma}$ with its diagonal elements $\lambda_i^{\alpha,\beta,\gamma}$,

$$\lambda_i^{\alpha,\beta,\gamma} = \begin{cases} \frac{\lambda_i + \alpha + \gamma}{K}, & \text{the } r \text{ positive eigenvalues with } \lambda_i > 0 \\ \frac{\alpha + \beta}{K}, & \text{otherwise} \end{cases}. \tag{7}$$

where K is a normalization constant given by

$$K = (d\alpha + (d - r)\beta + r\gamma + \text{tr}(\Xi_D)) / \text{tr}(\Xi_D). \tag{8}$$

2.2 Tensor Subspace Analysis

The natural representation of a facial image is matrix, which can also be regarded as the second order tensor X . Similarly, each projection vector \mathbf{a} can be expressed as a matrix \mathbf{A} . In tensor subspace analysis, we assume that the matrix \mathbf{A} can be expressed as a rank-1 matrix,

$$\mathbf{A} = u_1 \circ v_1. \tag{9}$$

Given the column projection matrix \mathbf{T}_c and the row projection matrix \mathbf{T}_r , we can derive a low rank tensor of the original tensor X using the N-mode multiplication,

$$Y = (X \times_1 T_r) \times_2 T_c = \mathbf{T}_r \mathbf{X} \mathbf{T}_c. \tag{10}$$

The target of tensor subspace analysis is to compute the column projection matrix \mathbf{T}_c and the row projection matrix \mathbf{T}_r . By far, a number of approaches, such as generalized low rank approximations of matrices (GLRAM) [9], bi-directional PCA (BDPCA) [11], and multilinear singular value decomposition [8], have been proposed to calculate these two projection matrices. Generally speaking, tensor subspace analysis is expected to have a good generalization performance for face recognition. However, during the calculation of the column or the row projection matrices, the tensor subspace analysis approaches usually neglects the spatial correlation in the other directions, correspondingly, and thus do not make full use of the available spatial correlation information in images.

2.3 Post-processed Subspace Analysis

Post-processed subspace analysis uses 2D-Gaussian filtering to make the projection vectors spatially smoothing [12, 13]. In face recognition, where the projection vector

can be transformed into a 2D image, Gaussian filtering is then used to post-process the projection vector and reduce noise. 2D-Gaussian function is defined as

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \tag{11}$$

where σ is the standard deviation. First, a 2D-Gaussian model M is defined according to the standard deviation $\sigma > 0$. The window size $[w, w]$ can then be determined as $w \approx 5 \times \sigma$, and the Gaussian model M is defined as the $w \times w$ truncation from the Gaussian kernel $G(x, y)$. We then calculate the norm of the projection vector $\|v_i\|_2 = \sqrt{v_i^T v_i}$, and map it into the corresponding projection image I_i . The filter M is used to smooth the projection image I_i ,

$$I'_i(x, y) = I(x, y) * M(x, y) \tag{12}$$

$I'_i(x, y)$ is transformed into a high dimensional vector v'_i by concatenating the rows of $I'_i(x, y)$ together. Finally we normalize v'_i using the norm of v_i

$$v''_i = \frac{\|v_i\|_2}{\sqrt{v_i^T v_i}} v'_i \tag{13}$$

and obtain the post-processed projection vector v''_2 .

2.4 Penalized Subspace Analysis

Cai's Penalized subspace analysis utilize the same procedure of regularized subspace analysis by using the Laplacian penalty to derive the regularized matrix [16]. The discrete approximation of the one-dimensional Laplacian function

$$D_j = \frac{1}{h_j^2} \begin{pmatrix} -1 & 1 & & & \\ 1 & -2 & 1 & & \\ & \cdot & \cdot & \cdot & \\ & & 1 & -2 & 1 \\ & & & 1 & -1 \end{pmatrix} \tag{14}$$

where h_j is the length of the one-dimensional signal. Further, the discrete version of the two-dimensional Laplacian is the $mn \times mn$ matrix,

$$\Delta = D_1 \otimes I_2 + I_1 \otimes D_2 \tag{15}$$

where I_1 is the $m \times m$ identity matrix, I_2 is the $n \times n$ identity matrix, and \otimes denotes the kronecker product.

Based on the the two-dimensional discrete Laplacian and the generalized graph embedding framework, the projection vectors of penalized subspace analysis can be calculated by maximizing the following criteria

$$w^* = \arg \max_w \frac{w^T X W X^T w}{(1-\alpha)w^T X D X^T w + \alpha J(w)}, \tag{16}$$

where J is the discrete Laplacian penalty function,

$$J(\mathbf{w}) = \mathbf{w}^T \Delta^T \Delta \mathbf{w} . \tag{17}$$

3 Penalized Subspace Analysis Using LOG and DOG Penalties

The target of Cai’s penalized subspace analysis is to make the projection vector spatially smooth by minimizing the Laplacian penalty function

$$J(f) = \int_{\Omega} \left[\sum_{j=1}^2 \frac{\partial^2 f}{\partial t^2} \right]^2 dt . \tag{18}$$

In this section, we will propose several alternative penalty functions for learning spatially smooth subspace. Taking into account the multi-scale characteristic of facial images, we propose a Laplacian of Gaussian (LOG) penalty function

$$J(f) = \int_{\Omega} \left[\sum_{j=1}^2 G(\sigma) * \frac{\partial^2 f}{\partial t^2} \right]^2 dt , \tag{19}$$

where $G(\sigma)$ is the Gaussian function with the variance σ , and $*$ denotes the convolution operator. Actually, the minimization of the square of the first order derivative would also be helpful in the learning of the spatially smooth subspace. We further propose a Derivative of Gaussian (DOG) penalty function

$$J(f) = \int_{\Omega} \left[\sum_{j=1}^2 G(\sigma) * \frac{\partial f}{\partial t} \right]^2 dt . \tag{20}$$

It should be noted that both the LOG and the DOG penalty functions have a variance parameter σ . Thus we could tune the σ value to meet the multi-scale property of the facial images. If the size of the facial image is high, we can use a large scale parameter σ to control the spatial smoothness.

After determining the variance σ of the Gaussian function, we can use a similar method to Cai’s penalized subspace to derive the discrete approximation of the LOG and DOG. For example, assuming $\sigma = 0.5$, the discrete approximation of the LOG function is represented as

$$D_j^{LOG} = \frac{1}{h_j^2} \begin{pmatrix} -0.4738 & 0.3903 & 0.1443 & 0.0116 & & & & & & \\ 0.3903 & -0.9364 & 0.3181 & 0.1441 & 0.0116 & & & & & \\ 0.1443 & 0.3181 & -0.9475 & 0.3181 & 0.1441 & 0.0116 & & & & \\ 0.0116 & 0.1441 & 0.3181 & -0.9475 & 0.3181 & 0.1441 & 0.0116 & & & \\ & & & & & & & & & \\ & & & 0.0116 & 0.1441 & 0.3181 & -0.9475 & 0.3181 & 0.1441 & 0.0116 \\ & & & & 0.0116 & 0.1441 & 0.3181 & -0.9475 & 0.3181 & 0.1443 \\ & & & & & 0.0116 & 0.1441 & 0.3181 & -0.9364 & 0.3903 \\ & & & & & & 0.0116 & 0.1443 & 0.3903 & -0.4738 \end{pmatrix} . \tag{21}$$

Further, the discrete version of the two-dimensional LOG function Δ_{LOG} and the DOG function Δ_{DOG} ,

$$\Delta_{LOG} = D_1^{LOG} \otimes I_2 + I_1 \otimes D_2^{LOG}, \Delta_{DOG} = D_1^{DOG} \otimes I_2 + I_1 \otimes D_2^{DOG} . \tag{22}$$

Finally, the criteria of penalized subspace analysis using the LOG and DOG penalties is defined as

$$\mathbf{w}^* = \arg \max_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{X} \mathbf{W} \mathbf{X}^T \mathbf{w}}{(1 - \alpha) \mathbf{w}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{w} + \alpha \mathbf{w}^T \Delta_{LOG}^T \Delta_{LOG} \mathbf{w}} , \tag{23}$$

$$\mathbf{w}^* = \arg \max_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{X} \mathbf{W} \mathbf{X}^T \mathbf{w}}{(1 - \alpha) \mathbf{w}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{w} + \alpha \mathbf{w}^T \Delta_{DOG}^T \Delta_{DOG} \mathbf{w}} . \tag{24}$$

4 Experimental Results and Discussion

In this section, we use two databases, the CMU PIE database [18] and the ORL face database (<http://www.cl.cam.ac.uk/Research/DTG/attarchive/facesataglance.html>), to evaluate the efficiency of the proposed penalty functions for spatially smoothing subspace face recognition methods. We implement two subspace analysis approaches, Fisherfaces and Laplacianfaces, and compare the recognition accuracy of the original methods and the corresponding penalized subspace analysis using the Laplacian, LOG, and DOG penalties.

4.1 Experiments on the ORL Database

The ORL face database is used to test the proposed spatially smoothing subspace face recognition methods. The ORL database contains 400 facial images with 10 images per individual. All the images are taken against a dark homogeneous background but vary in sampling time, light conditions, facial expressions, facial details (glasses/no glasses), scale and tilt. The size of these images is 112×92 . In our experiments, each facial image is manually aligned, cropped, and resized into a 32×32 image with 256 gray levels.

For the proposed method, we set the regularization parameter $\alpha = 0.1$, and the variance $\sigma = 0.5$. In our experiments, we separate the database into two subsets, the training subset and the test set, and use $\text{Tr}d/\text{T}et$ to denote that we randomly select d images of each individual for training and use the remained t images of each individual for testing. We run each face-recognition method ten times and calculate the average recognition rate to reduce the recognition rate variation caused by using different training and test set.

Table 1 lists the recognition rates of Fisherfaces, penalized Fisherfaces using the Laplacian (PF-L), LOG (PF-LOG), and DOG (PF-DOG) penalties. Two facts can be observed from this table. First, penalized Fisherfaces using both LOG and DOG penalties could achieve higher recognition accuracy than the original Fisherfaces method and the penalized Fisherfaces method with Laplacian penalty. Second, with the increasing of the training samples, because more stable diagonal matrix estimation could be achieved, the effectiveness of penalized Fisherfaces would be decreased.

Table 1. Recognition rates of Fisherfaces-based methods on ORL (mean±std. %)

Method	Tr2/Te8	Tr3/Te7	Tr4/Te6	Tr5/Te5
Fisherfaces	77.36±2.03	86.66±1.75	91.71±1.27	93.91±1.54
PF-L	85.31±2.17	92.25±1.29	95.31±0.92	97.00±1.03
PF-LOG	85.80±1.92	92.59±1.13	95.65±1.02	97.41±0.93
PF-DOG	86.41±2.13	92.48±1.36	96.04±1.11	97.56±1.17

Table 2 lists the average recognition rates of Laplacianfaces, penalized Laplacianfaces using the Laplacian, PL-LOG, and PL-DOG. For the Laplacianfaces-based methods, similar experimental results are obtained to the Fisherfaces-based methods. LOG and DOG penalties would also achieve higher recognition accuracy than the original Laplacianfaces method and the penalized Laplacianfaces method with Laplacian penalty. The experimental results further verify the effectiveness of the proposed two penalty functions.

Table 2. Recognition accuracy of Laplacianfaces-based methods on ORL (mean±std. %)

Method	Tr2/Te8	Tr3/Te7	Tr4/Te6	Tr5/Te5
Laplacianfaces	76.47±2.72	84.00±2.76	89.64±1.15	92.00±1.63
PL-L	84.78±1.89	92.23±1.58	95.10±1.59	96.84±1.03
PL-LOG	84.95±2.35	92.63±1.59	95.73±1.21	97.37±0.81
PL-DOG	86.01±2.34	92.50±1.59	95.58±1.15	97.50±1.08

4.2 Experiments on the CMU PIE Database

The CMU PIE face database contains 68 subjects captured under various pose, illumination, and expression. In our experiments, we choose the five near frontal poses (C05, C07, C09, C27, C29) with different illuminations and expressions to construct a face subset of 68×170 facial images. In our experiments, each facial image is manually aligned, cropped, and resized into a 32×32 image with 256 gray levels. We separate the database into two subsets, the training subset and the test set, and use Trd to denote that we randomly select d images of each individual for training and use the remained images of each individual for testing. We run each method ten times and calculate the average recognition rate to reduce the variation of recognition rate.

Table 1 lists the recognition rates of Fisherfaces, PF-L, PF-LOG, and PF-DOG. PF-LOG and PF-DOG could achieve higher recognition accuracy than the original Fisherfaces method, and the penalized Fisherfaces method with Laplacian penalty, which indicate the effectiveness of the proposed penalty functions.

Table 3. Recognition accuracy of Fisherfaces-based methods on CMU PIE (mean±std. %)

Method	Tr5	Tr10
Fisherfaces	48.05±1.34	64.20±0.71
PF-L	60.05±1.53	68.81±0.76
PF-LOG	61.22±1.61	70.15±0.78
PF-DOG	61.52±1.64	70.56±0.72

Table 4 presents the average recognition rates and standard deviations of Laplacianfaces, PL-L, PL-LOG, and PL-DOG. For the Laplacianfaces-based methods, PL-LOG and PL-DOG would also achieve higher average recognition accuracy than the original Laplacianfaces method, and penalized Laplacianfaces with Laplacian penalty.

Table 4. Recognition accuracy of Laplacianfaces-based methods on CMU PIE (mean \pm std. %)

Method	Tr5	Tr10
Laplacianfaces	48.45 \pm 1.60	55.05 \pm 0.77
PL-L	53.77 \pm 1.87	61.23 \pm 0.92
PL-LOG	54.87 \pm 1.74	62.61 \pm 0.75
PL-DOG	55.49\pm1.60	63.13\pm0.54

5 Conclusion

In this paper, we proposed two subspace face recognition approaches, spatially smooth subspace analysis using the LOG and DOG penalties. Both the LOG and the DOG penalty functions have a variance parameter σ which can be used to meet the multi-scale property of the facial images. Thus the proposed methods are more flexible in tuning the spatial smoothness. We use two databases, the CMU PIE database and the ORL face database to evaluate the proposed spatially smoothing subspace face recognition methods. Our experimental results show that the proposed methods are effective for face recognition, and achieve higher recognition accuracy than the original subspace analysis and Cai's penalized subspace analysis methods.

Acknowledgments. The work is partially supported by the NSFC fund under Contract Nos. 60620160097, 60571025, and 60872099, and the 863 fund under Contract Nos. 2006AA01Z193 and 2006AA01Z308.

References

1. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face Recognition: A Literature Survey. *ACM Computing Surveys* 35, 399–458 (2003)
2. Belhumeur, P., Hefanaha, J., Kriegman, D.: Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Trans. Pattern Analysis and Machine Intelligence* 19, 711–720 (1997)
3. He, X., Yan, S., Hu, Y., Niyogi, P., Zhang, H.J.: Face Recognition Using Laplacianfaces. *IEEE Trans. Pattern Analysis and Machine Intelligence* 27, 328–340 (2005)
4. Cevikalp, H., Neamtu, M., Wilkes, M., Barkana, A.: Discriminative common vectors for face recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence* 27, 4–13 (2005)
5. Lu, J., Plataniotis, K.N., Venetsanopoulos, A.N.: Regularization Studies of Linear Discriminant Analysis in Small Sample Size Scenarios with Application to Face Recognition. *Pattern Recognition Letters* 26, 181–191 (2005)
6. Zhao, W., Chellappa, R., Phillips, P.J.: Subspace linear discriminant analysis for face recognition. Technical Report CAR-TR-914, CS-TR-4009, University of Maryland (1999)

7. Dai, D.Q., Yuen, P.C.: Regularized Discriminant Analysis and Its Application to Face Recognition. *Pattern Recognition* 36, 845–847 (2003)
8. Lathauwer, L., Moor, B., Vandewalle, J.: A Multilinear Singular Value Decomposition. *J. Matrix Anal. Appl.* 21, 1253–1278 (2000)
9. Ye, J., Janardan, R., Li, Q.: Two-Dimensional Linear Discriminant Analysis. *NIPS* 17 (2004)
10. He, X., Cai, D., Niyogi, P.: Tensor Subspace Analysis. *NIPS* 18 (2005)
11. Zuo, W., Zhang, D., Wang, K.: Bi-Directional PCA with Assembled Matrix Distance Metric for Image Recognition. *IEEE Trans. Systems, Man, and Cybernetics, Part B* 36, 863–872 (2006)
12. Zuo, W., Wang, K., Zhang, D., Yang, J.: Regularization of LDA for Face Recognition: A Post-processing Method. In: *IEEE Int'l Workshop on Analysis and Modeling of Faces and Gestures*, pp. 377–391 (2005)
13. Hao, J., Zuo, W., Wang, K.: Theoretical Investigation on Post-Processed LDA for Face and Palmprint Recognition. In: *Int'l Conf. Computational Intelligence and Security*, pp. 301–305 (2007)
14. Wang, L., Zhang, Y., Feng, J.: On the Euclidean Distance of Images. *IEEE Trans. Pattern Analysis and Machine Intelligence* 27, 1334–1339 (2005)
15. Hastie, T., Buja, A., Tibshirani, R.: Penalized Discriminant Analysis. *Annals of Statistics* 23, 73–102 (1995)
16. Cai, D., He, X., Hu, Y., Han, J., Huang, T.: Learning a Spatially Smooth Subspace for Face Recognition. In: *Proc. 2007 IEEE Conf. Computer Vision and Pattern Recognition (CVPR 2007)* (2007)
17. Yan, S., Xu, D., Zhang, B., Zhang, H.J., Yang, Q., Lin, S.: Graph Embedding and Extension: A General Framework for Dimensionality Reduction. *IEEE Trans. Pattern Analysis and Machine Intelligence* 29, 40–51 (2007)
18. Sim, T., Baker, S., Bsat, M.: The CMU Pose, Illumination, and Expression Database. *IEEE Trans. Pattern Analysis and Machine Intelligence* 25, 1615–1618 (2003)

Nonnegative-Least-Square Classifier for Face Recognition

Nhat Vo, Bill Moran, and Subhash Challa

The University of Melbourne, VIC, Australia
n.vo@pgrad.unimelb.edu.au,
b.moran@ee.unimelb.edu.au,
subhash.challa@nicta.com.au

Abstract. In this paper, we propose a novel classification method, based on Nonnegative-Least-Square (NNLS) algorithm, for face recognition. Different from traditional classifiers, in our classifier, we consider each new sample (face) as a nonnegative linear combination of training samples (faces). By forcing the nonnegative constraint on linear coefficients, we obtain the nonnegative sparse representation that automatically discriminates between those classes present in the training set. Experimental results show the promising aspects of new classifier when comparing with the most popular classifiers such as Nearest Neighborhood (NN), Nearest Centroid (NC), and Nearest Subspace (NS) in terms of recognition accuracy, efficiency, and numerical stability. Eigenfaces, Fisherfaces, and Laplacianfaces are performed on Yale and ORL databases as feature extraction in these experiments.

Keywords: Face Recognition, Eigenfaces, Fisherfaces, and Nonnegative-Least-Square.

1 Introduction

With the rapidly increasing demand on *Face Recognition* (FR) technology, it is not surprising to see an overwhelming amount of research publications on this topic in recent years. Principal component analysis (PCA) or Eigenfaces [1] and Linear discriminant analysis (LDA) [2] or Fisherfaces are the most popular subspace analysis approaches to learn the low-dimensional structure of high dimensional face data. PCA finds a set of representative projection vectors such that the projected samples retain most information about original samples, while LDA finds a set of vectors that maximizes Fisher Discriminant Criterion, i.e. maximizes the ratio

$$\frac{w^T S_b w}{w^T S_w w} \quad (1)$$

where S_b is the between-class scatter matrix, and S_w is the within-class scatter matrix. This ratio is maximized when the column vectors w of the projection matrix W are the eigenvectors of $S_w^{-1} S_b$. Unfortunately, in face recognition tasks, this method cannot be applied directly since the dimension of the sample space is

typically larger than the number of samples in the training set. As a consequence, S_w is singular. This problem is known as the “small sample size problem” [3]. A lot of methods have been proposed to solve this problem [2][4][5][6][7][8]. In [2], they proposed a two stage PCA+LDA method, also popularly known as the Fisherfaces method, in which PCA is first used for dimension reduction so as to make S_w nonsingular before the application of LDA. Some other approaches are [4], Direct-LDA [5], Null space based LDA (NLDA) [6][7]. Besides PCA and LDA, there is another linear subspace method called Laplacianfaces or Locality Preserving Projections (LPP)[9] that seeks to preserve the intrinsic geometry of the data and local structure. The objective function of LPP is to minimize the local quantity, i.e., the local scatter of the projected data. During classification stage in face recognition, there are many methods used to classify faces, including nearest neighbor (NN), nearest centroid (NC), nearest subspace (NS), neural networks, and so on. Among those methods, NN is the simplest yet most popular method for classification task. Related to NN, NC and NS also are used for classifier due to their simplicity and efficiency. In this paper, we propose a novel classification method, based on Nonnegative Least Square algorithm (NNLS), for face recognition. Different from traditional classifiers, in our classifier, we consider each new sample (face) as a nonnegative linear combination of training samples (faces). By forcing the nonnegative constraint on linear coefficients, we obtain the nonnegative sparse representation that automatically discriminates between those classes present in the training set. Eigenfaces [1], Fisherfaces [2], and Laplacianfaces [9] are the most popular ones and chosen as feature extraction module for experiments in this paper. The outline of this paper is as follows. In Section 2, a brief introduction of traditional classifiers are presented and detail of the proposed classifier is also described. In Section 3, experimental results are performed for Yale and ORL face databases to demonstrate the effectiveness and promise of our new classifier. Finally, conclusions are presented in Section 4.

2 Classifiers for Face Recognition

In the field of machine learning, the goal of classification is to classify patterns that have similar feature values, into same classes. Among those classifiers, nearest neighbor (NN), nearest centroid (NC), and nearest subspace (NS) are the most used ones due to their easy implementation and efficiency. In this section, we briefly introduction these three classifiers, then present our new classifier based on Nonnegative Least Square algorithm.

2.1 Nearest Neighbor (NN)

The Nearest Neighbor (NN) [10] classifier is a method for classifying a new input feature (or sample) upon training feature vectors (or samples). Given a training set of feature vectors (or samples) $\{y_1, y_2, \dots, y_N\}$ taken values in a metric space, e.g. \mathcal{R}^n , with a priori known class $\{l_1, l_2, \dots, l_N\}$ respectively, where

$l_i \in \{1, 2, \dots, C\}$. In the testing stage, given a new feature vector (or sample) y and a defined metric distance, e.g. Euclidian, the geometric distance is computed between the new input feature vector y and each feature vector y_i from the training set to decide the nearest neighbor which is the one with shortest distance. And the class label l_k of the nearest neighbor is now assigned to the new input sample, i.e.

$$k = \arg \min_i \|y - y_i\|_2 \tag{2}$$

2.2 Nearest Centroid (NC)

In NC, we assign the testing feature vector (or sample) y to class label j if the distance from y to the mean feature vector (or sample) of class j^{th} is minimum, i.e.

$$j = \arg \min_i \|A_i e_i - y\|_2 \tag{3}$$

where $i = 1..C$, A_i is a matrix whose columns are training feature vectors (or samples) of class i^{th} , $e_i = [1/N_i, 1/N_i, \dots, 1/N_i]^T \in \mathbb{R}^{N_i}$ and N_i is the number of training feature vectors (or samples) in class i^{th} .

2.3 Nearest Subspace (NS)

The idea behind this is that each new feature vector (or sample) is assumed to lie in the subspace spanned by training feature vectors (or samples) of a specific class. We assign the testing feature vector (or sample) y to class label j if the distance from y to the subspace spanned by training feature vectors (or samples) belongs to that class is minimum, i.e.

$$j = \arg \min_i \left\{ \min_{\alpha} \|A_i \alpha - y\| \right\} \tag{4}$$

where $i = 1..C$, and A_i is a matrix whose columns are training feature vectors (or samples) of class i^{th} .

2.4 Nonnegative-Least-Square (NNLS)

The method of least squares is used to solve systems which can be stated as:

$$\min_x \|Ex - f\|_2 \tag{5}$$

where E is an n -by- m matrix, f is a given n element vector and x is the m element solution vector. However, in many real-world problems, the underlying parameters represent quantities that can take on only nonnegative values. In such a case, Problem (5) must be modified to include nonnegativity constraints on x , leading to a problem called Nonnegative Least Squares (NNLS) [11], and formulated as

$$\min_x \|Ex - f\|_2 \text{ s.t. } x \geq 0 \tag{6}$$

The NNLS problem is fairly old and the algorithm of Lawson and Hanson [11] seems to be apparently the first and most efficient method for solving it

(this algorithm is available as the *lsqnonneg* procedure in MATLAB and is used in our experiments).

2.5 Nonnegative-Least-Square Classifier (NNLSC)

The idea of proposing NNLSC is based on two points as follow:

- Each new sample is a sparse linear combination of training samples. We use nonnegative constraint to find this sparse representation.
- We believe that the nonnegative constraint prevents the overmatching problem that happens in traditional classifiers such as NN, NC or NS.

From these two points, we propose a simple classifier based on NNLS and called NNLSC. To illustrate how NNLSC works, we randomly select 5 samples for each class from ORL database to create 200 sample training data set, and select one sample from 1st class for testing. In this example, we use Eigenfaces method to extract features, and the reduced dimension of feature vector is 199. Fig. 1a illustrates the nonnegative coefficient vector by NNLS for a test image from 1st class and Fig. 1b shows the reconstruction error on each class of this test image. As we can see, the reconstruction error on 1st class of this test image is minimum, which means that this test image belong to 1st class.

Table 1. Algorithm – Nonnegative-Least-Square Classifier (NNLSC)

<p>Algorithm – NNLSC</p> <p>INPUT</p> <ul style="list-style-type: none"> – Given training feature matrix $Y = [Y_1, Y_2, \dots, Y_C] \in \mathbb{R}^{k \times N}$ from training feature samples of C classes. These features are obtained by performing Eigenfaces, Fisherfaces, or Laplacianfaces. – Feature transformation matrix $W \in \mathbb{R}^{n \times k}$ (k is reduced dimension). – Given new input image $x \in \mathbb{R}^n$. <p>ALGORITHM</p> <ul style="list-style-type: none"> – Calculate projected new sample or feature vector $y = W^T x \in \mathbb{R}^k$ and use NNLSC to find nonnegative coefficient vector $\hat{\alpha}$ with objective function as: $\min_{\alpha} \ Y\alpha - y\ _2 \text{ s.t. } \alpha \geq 0 \tag{7}$ – Suppose that $\hat{\alpha} = [r_1^T, r_2^T, \dots, r_C^T]$, we calculate reconstruction error on each class as $e_i = \ y - Y_i r_i\ _2$ with $i = 1..C$. <p>OUTPUT : x belongs to class j^{th}, where $j = \arg \min_i (e_i)$.</p>

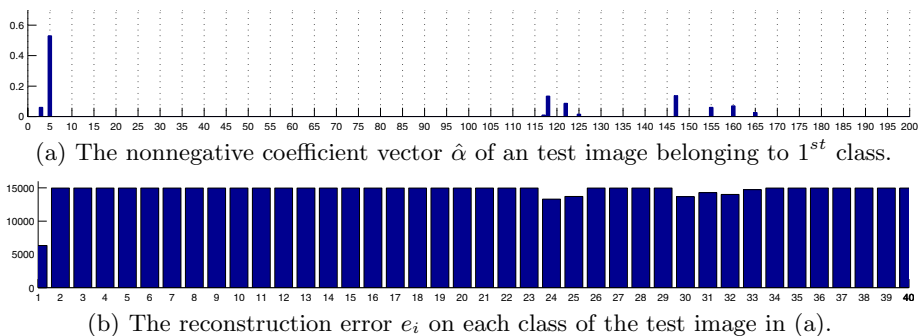


Fig. 1. Example of performing NNLSC on ORL database with Eigenfaces method

3 Experiments on Face Databases

This section evaluates the performance of Eigenfaces [1], Fisherfaces [2], Laplacianfaces [9] with NN, NC, NS and NNLSC classifiers using Yale and ORL databases. Some sample images from Yale and ORL databases are shown in Fig. 2 and Fig. 3. The Euclidean metric is used as our distance measure for all experiments. For those linear subspace methods such as Eigenfaces, Fisherfaces, and Laplacianfaces, the recognition process basically has three steps. First, we calculate the face subspace from the training set of face images; then the new face image to be identified is projected into low-dimensional subspace. Finally, the new face image is identified by using NN, NC, NS or NNLSC. It should be noted that in PCA, an upper bound on the dimension of the reduced space is $N - 1$, where N is the number of training samples, while the maximum reduced dimension of LDA is $C - 1$, where C is the number of classes or subjects. Though LPP is considered as unsupervised techniques, in our the experiment we adapt the supervised versions of these algorithms, details can be found in [9]. The reason we use the supervised versions of LPP for comparison is that LPP’s performance in supervised mode is better than that in the unsupervised mode.



Fig. 2. Ten images of a person with different facial expressions or configurations from Yale face database



Fig. 3. Twenty sample images of two people taken at different constraints (pose, lighting, ...) from ORL face database

3.1 Yale Face Database

The Yale face Database contains 165 grayscale images of size 100×100 in GIF format of 15 individuals. There are 11 images per subject, one per different facial expression or configuration: center-light, w/glasses, happy, left-light, w/no glasses, normal, right-light, sad, sleepy, surprised, and wink. A random subset with k ($k = 5, 7, 9$) images per individual was taken with labels to form the training set. The rest of the database was considered to be the testing set. 10 times of random selection for training examples were performed and the average recognition result was recorded. We tested the recognition rates with different number of training samples and show the best results obtained by Eigenfaces [1], Fisherfaces [2], and Laplacianfaces [9], with NN, NC, NS, and>NNLSC in Table 2, Table 3, and Table 4. The values in parentheses denote the dimension of feature vectors for the best recognition accuracy.

Table 2. Comparison of the top recognition accuracy (%) on Yale database with Eigenfaces method

k	NN	NC	NS	NNLSC
5	58.48% (32)	42.12% (39)	63.03% (23)	64.21% (26)
7	57.96% (46)	49.24% (32)	62.88% (66)	65.15% (39)
9	60.61% (8)	55.05% (66)	66.67% (60)	69.17% (39)

Table 3. Comparison of the top recognition accuracy (%) on Yale database with Fisherfaces method

k	NN	NC	NS	NNLSC
5	56.36% (10)	56.97% (10)	59.09% (10)	61.82% (10)
7	58.33% (10)	58.33% (10)	67.42% (10)	69.70% (10)
9	67.17% (10)	68.01% (10)	59.09% (10)	71.21% (10)

Table 4. Comparison of the top recognition accuracy (%) on Yale database with Laplacianfaces method

k	NN	NC	NS	NNLSC
5	67.03% (33)	65.82% (10)	66.36% (14)	69.55% (10)
7	70.91% (16)	69.09% (10)	70.68% (46)	76.14% (10)
9	70.00% (35)	69.09% (8)	70.00% (71)	75.09% (29)

3.2 ORL Face Database

In the ORL database (<http://www.cam-orl.co.uk>), there are ten different images of size 112×92 of each of 40 distinct subjects. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). A similar protocol of experiment is performed as that of the Yale database. Comparison of the top recognition accuracy (%) on ORL database can be seen in Table 5, Table 6, and Table 7.

Table 5. Comparison of the top recognition accuracy (%) on ORL database with Eigenfaces method

k	NN	NC	NS	NNLSC
3	88.71% (118)	80.96% (112)	88.21% (72)	91.60% (66)
4	92.21% (43)	84.75% (158)	92.67% (38)	95.03% (86)
5	93.30% (184)	84.95% (140)	93.85% (56)	95.75% (51)

Table 6. Comparison of the top recognition accuracy (%) on ORL database with Fisherfaces method

k	NN	NC	NS	NNLSC
3	88.71% (39)	88.79% (39)	89.18% (39)	91.50% (39)
4	92.92% (37)	92.92% (38)	91.67% (37)	94.54% (39)
5	93.33% (39)	93.67% (39)	94.50% (39)	95.83% (34)

Table 7. Comparison of the top recognition accuracy (%) on ORL database with Laplacianfaces method

k	NN	NC	NS	NNLSC
3	87.79% (39)	87.79% (39)	86.21% (94)	88.21% (67)
4	91.33% (39)	91.33% (39)	91.00% (138)	92.75% (142)
5	93.67% (40)	93.00% (39)	93.00% (120)	94.33% (188)

3.3 Result Analysis

Some observations from experimental results can be summarized as follow:

- In general, we can see that our proposed classifier NNLSC outperforms the other popular classifiers in term of recognition accuracy on these experiments.
- NN and NS are comparable as the second best method in both Yale and ORL database.
- In term of time complexity, the ranking order of these classifiers is NN, NC, NNLSC and NS. This means that the NNLSC runs faster than NS.

4 Conclusion

In this paper, we have proposed a simple classifier that exploits the sparse representation and nonnegative linear combination of face data representation. The approach gives very promising results on standard databases compared with the other traditional classifiers. While the current work focuses more on experimental and intuitive observations, a theoretical investigation need to be done to support the idea in future work.

References

1. Turk, M., Pentland, A.: Eigenfaces for recognition. *Journal of Cognitive Neuroscience* 3(1), 71–86 (1991)
2. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(7), 711–720 (1997)
3. Fukunaga, K.: Introduction to statistical pattern recognition. Academic Press Professional, Inc., San Diego (1990)
4. Zhao, W., Chellappa, R., Krishnaswamy, A.: Discriminant analysis of principal components for face recognition. In: *FG 1998: Proceedings of the 3rd. International Conference on Face & Gesture Recognition*, p. 336. IEEE Computer Society, Washington (1998)
5. Yu, H., Yang, J.: A direct lda algorithm for high-dimensional data - with application to face recognition. *Pattern Recognition* 34(10), 2067–2070 (2001)
6. Chen, L.F., Liao, H.Y.M., Ko, M.T., Lin, J.C., Yu, G.J.: A new lda-based face recognition system which can solve the small sample size problem. *Pattern Recognition* 33(10), 1713–1726 (2000)
7. Huang, R., Liu, Q., Lu, H., Ma, S.: Solving the small sample size problem of lda. In: *ICPR 2002: Proceedings of the 16 th International Conference on Pattern Recognition (ICPR 2002)*, vol. 3, p. 30029. IEEE Computer Society, Washington (2002)
8. Cevikalp, H., Neamtu, M., Wilkes, M., Barkana, A.: Discriminative common vectors for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(1), 4–13 (2005)
9. He, X., Yan, S., Hu, Y., Niyogi, P., Zhang, H.J.: Face recognition using laplacianfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(3), 328–340 (2005)
10. Cover, T., Hart, P.: Nearest neighbor pattern classification. *IEEE Transactions on Information Theory* 13(1), 21–27 (1967)
11. Lawson, C.L., Hanson, R.J.: Solving least squares problems. Prentice-Hall Series in Automatic Computation. Prentice-Hall, Englewood Cliffs (1974)

A Novel Model for Recognition of Compounding Nouns in English and Chinese

Lishu Li¹, Jiawei Chen¹, Qinghua Chen,¹ and Fukang Fang²

¹ Department of Systems Science, Beijing Normal University, Beijing 100875, China
² Institute of Non-equilibrium Systems,
Beijing Normal University, Beijing 100875, China
chenjiawei@bnu.edu.cn

Abstract. Compounds are very common in many kinds of language. Most of the research in this field is from the view of morphology, while artificial neural network is seldom concerned. Based on Hopfield model, we create a novel neural network to simulate the recognition process of compounds in English and Chinese. Our model is composed of two layers: abstraction layer and recognition layer. The first layer can extract the common features of the training samples and represent it as a new attractor, which can be transferred into the next layer. This step imitates morpheme abstraction of compounds. Recognition layer is constructed as an improved Hopfield network, in which two existing attractors can merge into a new one. This step reflects the cognition of a new compound when all the morphemes are memorized. One specific example ‘*raincoat*’ is demonstrated, and the results provide strong evidence to our model.

Keywords: Compounding nouns, Neural network, Hopfield model, Attractor.

1 Introduction

Compounding is a very important word formation in many kinds of language and the study of compounds has formed a complete system. Most of research is focused on morphology, which is mainly manipulated through corpus analysis or behavioral studies. Gary Libben’ work shows that both semantically transparent compounds and semantically opaque compounds show morphological constituency. The semantic transparency of the morphological head was found to play a significant role in overall lexical decision latencies, in patterns of decomposition, and in the effects of stimulus repetition within the experiment [1]. Todd R. Haskell proposed a new account in which the acceptability of modifiers is determined by a constraint satisfaction process modulated by semantic, phonological, and other factors [2]. Elena Nicoladis’ research results demonstrated that children’s knowledge of the meaning of compound nouns is still developing in the preschool years [3]. Some scholars apply biological method to this problem.

I. Cummings measures eye-movements during reading and found that morphological information becomes available earlier than semantic information during the processing of compounds [4]. B. J. Juhasz explored the role of semantic transparency for English compound words, and the analysis of gaze durations revealed that transparency did not interact with lexeme frequency, suggesting decomposition occurs for both transparent and opaque compounds [5]. There is also some investigation concerned bilingual study. Nivja H. de Jong uses the association between various measures of the morphological family and decision latencies to reveal the way in which the components of Dutch and English compounds are processed [6]. Elena Nicoladis explores the cues used in acquisition of two semantically similar structures that are ordered differently in French and English: adjectival phrases and compound nouns [7].

All of the research has obtained much achievement, while they mainly manipulate from the angle of morphology or behavioral experiment, but the neural mechanism of these processes are not clear. Since the cognition process is complicated, it is necessary for us to explore the inner kernel mechanism considering some real neural functions. Over the past three decades, artificial neural networks (ANNs), which are non-linear mapping structures based on the function of human brain, have been applied widely in computational neuroscience. Among those models, associative memory neural network is typical, which suggests that memories are represented as stable network activity states called attractors. When a stimulus pattern is presented to the system, the network dynamics are drawn toward the attractor that corresponds to the memory associated with that stimulus [8,9,10]. Some attractor networks have been created to study the problems about language learning [11,12]. In this paper, we bring forward a new kind of ANN based on Hopfield network, which is a classical associative memory network, to achieve the recognition of compound in Chinese and English.

We argue that the recognition process is divided into two steps. First, we learn each morpheme's meaning of the compound. We can accomplish this by abstracting each sense from many memorized compounding words which include the same morpheme. This step can be interpreted as a new attractor's formation. Once achieving the first approach, we can guess the compound's meaning by combining each constituent's sense. In this process we suppose that two existing attractor can merge into a new one. We also guess that both outcomes in the two steps can be interpreted as emergence, because they generate new attractor respectively.

This paper is organized into 4 sections. Section 1 is brief introduction referring to recent research in compound recognition. In the next section, we first review the necessary backgrounds of Hopfield network, then our model is proposed and the details about the learning rules is interpreted particularly. The simulation result is shown in section 3. In the final section we conclude our idea and discuss further work.

2 Model

2.1 Architecture of Network

There are two layers in our model: abstraction layer and recognition layer, and the neuron numbers in the two layers are equal. Each neuron in abstraction layer is associated with one unique neuron in recognition layer. The generation in the first layer will be transferred to the next and the second layer will perform the final output. The architecture of our model is illustrated in Fig. 1. The learning rules in each layer is different, which we will explain amply in the next two subsection.

The structure of each layer is the same as classical Hopfield neural network, which consists of N fully connected binary neurons [8,13]. Each neuron i has two states: $s_i = 0$ (not firing) and $s_i = 1$ (firing). When neuron i has a connection made to it from neuron j , the strength of connection is defined as w_{ij} (Nonconnected neurons have $w_{ij} \equiv 0$). The instantaneous state of the system is specified by listing the N values of s_i , so it is represented by a binary word of N bits.

The state changes in time according to the following algorithm. For each neuron i , there is a fixed threshold θ_i , neuron i readjusts its state randomly in time but with a rate w_{ij} , setting as

$$s_i = \begin{cases} 1 & \text{if } \sum_{i \neq j} w_{ij} s_j > \theta_i \\ 0 & \text{if } \sum_{i \neq j} w_{ij} s_j < \theta_i \end{cases} \quad (1)$$

$\sum_{i \neq j} w_{ij} s_j$ is the net input to neuron i . The input to a particular neuron arises from the current leaks of the synapses to that neuron, which influence the cell mean potential. The synapses are activated by arriving action potentials. Thus each neuron randomly and asynchronously evaluates whether it is above or below threshold and readjusts accordingly.

2.2 Learning Rule in Abstraction Layer

In the former theory of neural networks the weight w_{ij} is considered as a parameter that can be adjusted so as to optimize the performance of a network for a given task. In our model, we assume that the weight will be updated according to Hebbian learning rule [14], i.e. the network learns by strengthening connection weights between neurons activated at the same time. It can be written as following:

$$\Delta w_{ij} = \begin{cases} \eta \cdot w_{ij} - d, & \text{if } S_i = 1, S_i = 1 \\ -\eta \cdot w_{ij} - d, & \text{if } S_i = 1, S_i = 0 \\ -\eta \cdot w_{ij} - d, & \text{if } S_i = 0, S_i = 1 \\ -d, & \text{if } S_i = 0, S_i = 0 \end{cases} \quad (2)$$

here, $0 < \eta < 1$ is a small constant called learning rate. The parameter d is a small positive constant that describes the rate by which w_{ij} decays back to zero

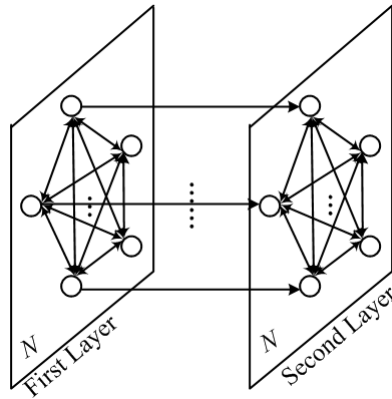


Fig. 1. The structure of model: abstraction layer (first layer) and recognition layer(second layer). Neurons in each layer are fully connected and they have two states: ‘0’ and ‘1’. The two layers have equivalent N neurons and the connections of neurons between layers are one-to-one.

in the absence of stimulation. Of course, equation (2) is just one of the possible forms to specify rules for the growth and decay of the weights, and there are some difference with the other forms of Hebbian rule [15].

From the formula (2) we can see that synaptic efficacy w_{ij} would grow without limit if the same potentiating stimulus is applied over and over again. A saturation of the weights should be consider. On the other hand, the synaptic efficacy w_{ij} should be non-negative. These two restrictions can be achieved by setting:

$$w_{ij}(t + 1) = \begin{cases} 1, & \text{if } w_{ij}(t + 1) > 1 \\ w_{ij}(t + 1), & \text{if } 0 \leq w_{ij}(t + 1) \leq 1 \\ 0, & \text{if } w_{ij}(t + 1) < 1 \end{cases} \quad (3)$$

2.3 Learning Rule in Recognition Layer

The learning rule in recognition performs according to Hebbian rule, which is an instance of an unsupervised learning procedure. In Hebbian learning, weights between learning neurons are adjusted so that each weight better represents the relationship between the neurons. Neurons which tend to be positive or negative at the same time will have strong positive weights while those which tend to be opposite will have strong negative weights. Neurons that are uncorrelated will have weights near zero. For example, if two neurons A and B are often simultaneously active, Hebbian learning will increase the connection strength between the two so that excitation of either one tends to cause excitation of the other. On the other hand, if neurons A and C were of opposite activations at all times, then Hebbian learning would gradually decrease the connection in between below zero so that an excited A or C would inhibit the other.

Formally, Hebb's rule for the modification of a weight w_{ij} from neuron i to j with a learning rate η is defined as

$$\Delta w_{ij} = \eta s_i s_j \quad (4)$$

Hebbian learning has four features interesting to the cognitive scientist: first it is unsupervised; second it is a local learning rule, meaning that it can be applied to a network in parallel; third it is simple and therefore requires very little computation; fourth it is biologically plausible.

The connectivity w_{ij} in traditional Hopfield's model is defined as

$$w_{ij} = \sum_{\mu=1}^L \xi_i^\mu \xi_j^\mu \quad (5)$$

It implies that all the information about the patterns to be memorized has been captured in the network. In our model, we use an improved form, which represents an optimal learning rule for associative memory networks [16,17].

$$w_{ij} = \frac{1}{Np(1-p)} \sum_{\mu=1}^L (\xi_i^\mu - p)(\xi_j^\mu - p) \quad (6)$$

ξ_i^μ ($= 1, \dots, L$) denote patterns to be memorized, L is the number of patterns. The variable p represents the mean level of activity of the network for L patterns.

3 Simulation

Using our model, we will demonstrate the cognition process of a English compound — 'raincoat', whose Chinese meaning is '雨衣'. 'rain' and 'coat' are two nouns, which means '雨' and '衣' respectively in Chinese. In our simulation, we suppose these two words are unknown initially, while the network can learn them by itself after being trained with some other compounding nouns which contain the constituents — 'rain' or 'coat'. Since the network has memorized the two words' English and Chinese meaning, the new compound 'raincoat' is input to the network to test whether it can be recognized.

There are two groups of neurons in our model, denoting $G1$, $G2$. Neurons in $G1$ are laid to a 16×68 matrix, representing English compounding nouns. $G2$ consists of 512 neurons, which are laid to a 16×32 matrix, to express the corresponding meaning of Chinese. The photographed letter or character is decomposed into pixels, and the value at each pixel corresponds to the value of a neuron in the network.

Our model implements three steps as follow. First, we train the network with the seven particular samples, which share the same part of 'rain' and its corresponding Chinese meaning '雨'. The samples are displayed in Fig.2. The training will not stop until the network can produce a stable output. Next, the output in the first layer is considered to be an input into the next layer as a pattern to



Fig. 2. Seven samples to be trained. All of the samples share the same part of ‘rain’ and its corresponding Chinese meaning.

be memorized. Since the two layers have equivalent neurons, the value of each neuron in abstraction layer is passed to its counterpoint in recognition layer. Finally, we present a new pattern ‘raincoat’ into the recognition layer, where the attractors are already in existence, to test whether the network can respond as ‘雨衣’ exactly. The values of parameters are set as: $\eta = 0.09$, $d = 0.5$, $\theta = 50$. The network is trained 60 times with samples selected from the training sample set arbitrarily and the original weights are set randomly. We display the simulation result step by step. The training outcome in abstraction layer is demonstrated in Fig. 3. We can find that as training time increases, the different parts of the samples vanish gradually, and the common feature, i. e. ‘rain’ and ‘雨’, are preserved. After training 60 times, the model can come out a steady output, which means a new attractor has engendered in the network.

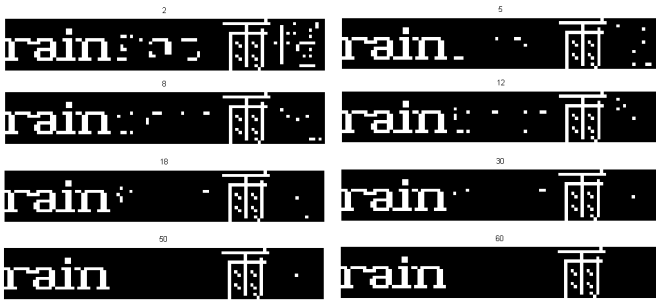


Fig. 3. The output of abstraction layer. As training proceeds, the discrepancy in the samples dies down gradually, and the uniform part are preserved. After training 60 times, the model can produce a stable output, which means a new attractor is generated.

As soon as the novel attractor is formed, it can be transferred to the second layer automatically. In the recognition layer, there is already an attractor — ‘coat’ memorized in our model, which is shown in Fig. 4, with implication that there are two patterns in the recognition layer in total. (The formation of this attractor is the same as ‘rain’, i. e., we can abstract this meaning through many compounds including morpheme ‘coat’. For simplification, we don’t describe the



Fig. 4. The other attractor memorized in recognition layer — ‘coat’ and its corresponding Chinese meaning

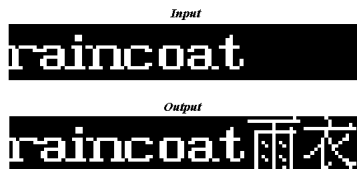


Fig. 5. The input and its final output of recognition layer. the unlearned compound ‘raincoat’ is not memorized in the network, but it has been recognized correctly. This result implies a new attractor is produced again.

details here.) ‘*raincoat*’ is a totally unconversant sample for the network, and the network didn’t store this pattern. According to traditional Hopfield network’s theory, it can’t produce a stable or meaningful output. But our simulation result make a conflict with this conclusion. The input and final output is shown in Fig. 5. We can see ‘*raincoat*’ can be recognized by the network, and the Chinese meaning ‘雨衣’ is correct, which implies that the model can generate a new attractor by itself. Besides, the new attractor integrates the information of the former attractors and stimulates the inactive part again, so we suppose it is the mergence of the two existing attractors. The simulation result can interpret the recognition process of compounding nouns: extracting each uniform constituent from concrete compounds which contain the same morpheme, then integrating all of the constituents and comprehending the new word.

4 Discussion

In this paper, we bring forward a novel neural network, which can simulate the recognition process of compounding nouns. In this model, especially in the abstraction layer, learning rule plays a key role. We consider two broad classes of operations dominate the approach: hypothesis elimination and associative learning. On the one hand, the features that all samples cover are called essential features, and the connections between neurons which represent the essential features will be strengthened during the training process. Associative learning works. On the other hand, the connections between individual features and that between individual features and essential features become weaker and weaker gradually. Hypothesis elimination works. These two operations can be precisely described by Hebbian learning rule. So Hebbian learning rule may be the neural mechanism of compound cognition in certain condition.

From the viewpoint of complexity theory, we can also find some evidence supporting our idea. In the recognition layer, the model behaves as an associative

memory when the state space flow generated by the algorithm is characterized by a set of stable fixed points. If these stable points describe a simple flow in which nearby points in state space tend to remain close during the flow, then initial states that are close to a particular stable state and far from all others will tend to terminate in that nearby stable state. As we mentioned above, each attractor has its own basin of attraction, and sometimes they have overlap partly. When the initial state is located in such region, which fixed point might it reach at last? The final result of our simulation throws light on this problem: it will arrive at neither of the fixed points but generate a new attractor instead. We also guess that both outputs in the two layers can be interpreted as emergence, because they generate new attractor respectively.

There are still further steps in our work. Our model can demonstrate the cognition of compounds whose meaning can be inferred directly only by integrating each morpheme's sense. It is well known that the classifications of compounds according to the classes of words are diversiform, besides, the relations between morpheme and compounds are very complicated. Compounds may be distinguished from free phrases on phonological, semantic, grammatical and orthographical features. There are still some compounds including two interpretable parses (e.g., '*rainbow*'), yet they has extended the integrated meaning of each morpheme. The facts suggest that morphological parsing does not simply divide a word into its constituents and combine the meanings easily, in this case, more kinds of factors and more complex mechanisms should be considered.

Acknowledgement. This work is supported by NSFC under the grant No. 60534080 and 60774085.

References

1. Libben, G., Compound, F.: The role of semantic transparency and morphological headedness. *Brain and Language* 84, 50–64 (2003)
2. Haskell, T.R.: Language learning and innateness: Some implications of Compounds Research. *Cognitive Psychology* 47, 119–163 (2003)
3. Nicoladis, E.: What compound nouns mean to preschool children. *Brain and Language* 84, 38–49 (2003)
4. Cunnings, I.: The time-course of morphological constraints: Evidence from eye-movements during reading. *Cognition* 104, 476–494 (2007)
5. Juhasz, B.J.: The influence of semantic transparency on eye movements during English compound word recognition. *Eye Movements: A Window on Mind and Brain*, 373–389 (2007)
6. Jong, N.H.D.: The Processing and Representation of Dutch and English Compounds: Peripheral Morphological and Central Orthographic Effects. *Brain and Language* 81, 555–567 (2002)
7. Nicoladis, E.: The Cues That Children Use in Acquiring Adjectival Phrases and Compound Nouns: Evidence from Bilingual Children. *Brain and Language* 81, 635–648 (2002)
8. Hopfield, J.: Neural networks and physical systems with emergent collective computational. *Proc. Natl. Acad. Sci. USA* 79, 2554–2558 (1982)

9. Amit, D.: The Hebbian paradigm reintegrated: local reverberations as internal representations. *Behav. Brain Sci.* 18, 617–657 (1995)
10. Brunel, N.: Network models of memory, In *Methods and Models in Neurophysics*. In: Volume Session LXXX: Lecture Notes of the Les Houches Summer School 2003, pp. 407–476 (2005)
11. Harm, M.W.: Phonology: Reading Acquisition, and Dyslexia: Insights from Connectionist Models. *Psychological Review* 106, 491–528 (1999)
12. Hinton, G.E.: Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review* 98, 74–95 (1991)
13. Hopfield, J.: Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of National Academic Science, USA* 81, 3088–3092 (1984)
14. Hebb, D.O.: *The Organization of Behavior*. Wiley Press, New York (1949)
15. Gerstner, W., Kistler, W.M.: *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge University Press, Cambridge (2002)
16. Dayan, P.: Optimizing synaptic learning rules in linear associative memories. *Biol. Cyber.* 65, 253–265 (1991)
17. Palm, G.: *Models of neural networks III*. Springer, Berlin (1996)

Orthogonal Quadratic Discriminant Functions for Face Recognition

Suicheng Gu, Ying Tan, and Xingui He

Key Laboratory of Machine Perception(MOE), Peking University;
Department of Machine Intelligence, School of Electronics Engineering and Computer
Science, Peking University, Beijing, 100871, P.R. China
ytan@pku.edu.cn

Abstract. Small sample size (SSS) problem is usually a limit to the robustness of learning methods in face recognition. Especially in the quadratic discriminant functions (QDF), too many parameters need to be estimated and covariance matrix of a class is usually singular. In order to overcome the SSS problems, we proposed a novel approach called orthogonal quadratic discriminant functions (OQDF). The OQDF assumes probability distribution functions of each two classes of face images have a uniform shape. Then, three OQDF models are developed. The Laplacian smoothing transform (LST) and Fisher's linear discriminant (FLD) are employed to preprocess the face images for the OQDF classifier. Finally, we evaluate our proposed algorithms on two face databases, ORL and Yale.

Keywords: Orthogonal quadratic discriminant functions (OQDF), modified quadratic discriminant function (MQDF), small sample size (SSS), face recognition (FR), Laplacian Smoothing Transform(LST), Fisher's linear discriminant(FLD).

1 Introduction

Face recognition has been an active research point in pattern recognition field for several decades. Statistical approaches, neural network approaches [1], support vector machine (SVM) [2] have all been well studied on this problem. Among all these approaches, the statistical approaches are always favored in practical applications, due to their robust characteristic and simple training schemes.

Quadratic discriminant function (QDF) is one of the most commonly used nonlinear techniques for pattern classification. In the QDF framework, the class conditional distribution is assumed to be Gaussian, however, with an allowance for different covariance matrices. Due to the fact that many free parameters are to be estimated (C covariance matrices, where C denotes the number of classes), QDF is susceptible to the so-called small sample size (SSS) problem in which the number of training samples is smaller or comparable to the dimensionality of the sample space.

The modified QDF (MQDF) proposed by Kimura *et al.* [3] aims to improve the computational efficiency and classification performance of the QDF via eigenvalue smoothing. It has been extremely successful and widely used for handwritten character recognition [4]. Alternatively, the regularized discriminant analysis (RDA) of Friedman [5] improves the performance of QDF by adding a small multiple of the identity matrix to the covariance matrix.

Due to the SSS problem to the QDF approaches, few QDF based algorithms were implemented for face recognition. But, the regularized direct QDA (RDQDA) [6] and Kernel quadratic discriminant analysis (KRQDA) [7] are two exceptions. The RDQDA employed the D-LDA [8] to reduce dimension first, then used the RDA to train. The KRQDA replaced D-LDA with kernel machine. However, these two approaches have been implemented based on the dimension reduction techniques. Both of them didn't provide the advantages over other classifiers. Actually, their advantages might due to efficient feature extraction strategies.

In this paper, we assume probability distribution functions of each two classes of face images have a uniform shape. An efficient strategy can be implemented to accelerate the computation. The quadratic discriminant functions under this assumption is called orthogonal quadratic discriminant function (OQDF). Unlike the QDF, regularized QDF and MQDF, the OQDF set the covariance matrices with same eigenvalues.

To extract features of the face images, we just select the two step feature extraction model, i.e., Laplacian smoothing transform (LST) [9] plus Fisher's Linear Discriminant [10].

The rest of this paper is organized as follows. The QDF and MQDF are reviewed briefly in section 2. In section 3, the proposed OQDF is deduced. Section 4 presents the experimental results. Finally, a conclusion is given in Section 5.

2 QDF and Modified QDF

Based on Bayesian decision rule, used to classify the input pattern to the class as maximum a posteriori (MAP) probability, the quadratic discriminant function (QDF) is obtained under the assumption of multivariate Gaussian density for each class. The MQDF proposed by Kimura *et al.* [3] makes a modification to the QDF by K-L transform and smoothing the minor eigenvalues to improve its computation efficiency and classification performance.

2.1 QDF

Consider a C class problem. Let $X = (x_1, \dots, x_N)$ denote N training samples, each x_i with dimension d . The QDF is obtained by

$$g_0(x, \omega_i) = (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i) + \log |\Sigma_i| \tag{1}$$

where μ_i and Σ_i denote the mean vector and the covariance matrix of class ω_i . The QDF can be used as a distance metric in the sense that the class of minimum distance is assigned to the input pattern.

By K-L transform, the covariance matrix can be diagonalized as

$$\Sigma_i = \Phi_i \Lambda_i \Phi_i^T, \tag{2}$$

where $\Lambda_i = \text{diag}[\lambda_{i1}, \dots, \lambda_{id}]$ with $\lambda_{ij}, j = 1, \dots, d$, being the eigenvalues (ordered in decreasing order) of Σ_i and $\Phi_i = [\phi_{i1}, \dots, \phi_{id}]$ with $\phi_{ij}, j = 1, \dots, d$, being the ordered eigenvectors. Φ_i is orthogonal (unitary) such that $\Phi_i^T \Phi_i = I$.

According to Eq. (2), the QDF can be rewritten in the form of eigenvectors and eigenvalues as

$$\begin{aligned} g_0(x, \omega_i) &= [\Phi_i^T(x - \mu_i)]^T \Lambda_i^{-1} [\Phi_i^T(x - \mu_i)] + \log |\Lambda_i| \\ &= \sum_{j=1}^d \frac{1}{\lambda_{ij}} [\phi_{ij}^T(x - \mu_i)]^2 + \sum_{j=1}^d \log \lambda_{ij}. \end{aligned} \tag{3}$$

2.2 MQDF

By replacing the minor eigenvalues with a constant σ_i in Eq. (3), the MQDF is obtained as

$$\begin{aligned} g_m(x, \omega_i) & \tag{4} \\ &= \sum_{j=1}^k \frac{1}{\lambda_{ij}} [\phi_{ij}^T(x - \mu_i)]^2 + \sum_{j=1}^k \log \lambda_{ij} + \sum_{j=k+1}^d \frac{1}{\sigma_i} [\phi_{ij}^T(x - \mu_i)]^2 + (d - k) \log \sigma_i \\ &= \frac{1}{\sigma_i} \|x - \mu_i\|^2 + \sum_{j=1}^k \left(\frac{1}{\lambda_{ij}} - \frac{1}{\sigma_i} \right) [\phi_{ij}^T(x - \mu_i)]^2 + \sum_{j=1}^k \log \lambda_{ij} + (d - k) \log \sigma_i, \end{aligned}$$

where k denotes the number of principal eigenvectors. Eq. (4) utilizes the invariance of Euclidean distance:

$$d_E(x, \omega_i) = \|x - \mu_i\|^2 = \sum_{j=1}^d [\phi_{ij}^T(x - \mu_i)]^2. \tag{5}$$

Since the training of the QDF classifier always underestimate the patterns eigenvalues by limited sample set, the minor eigenvalues become some kind of unstable noises and affect the robustness of classifier. By smoothing them in the MQDF classifier, not only the classification performance is improved, but also the computation time and storage for the parameters are saved.

The parameter σ_i can be set to a class-independent constant as proposed by Kimura et al. [3] (called as MQDF2). Moghaddam and Pentland [11] used a class-dependent constant calculated by the average of minor eigenvalues (called as MQDF3):

$$\sigma_i = \frac{\text{tr}(\Sigma_i) - \sum_{j=1}^k \lambda_{ij}}{d - k} = \frac{1}{d - k} \sum_{j=k+1}^d \lambda_{ij}, \tag{6}$$

where $\text{tr}(\Sigma_i)$ denotes the trace of covariance matrix. [4] and [12] found that the performance is superior when setting the constant class independent rather than class-dependent.

The QDF can be also combined with the regularized discriminant analysis (RDA) by interpolating the covariance matrices and then replacing the minor eigenvalues with the average in the complement subspace. We call the QDF combine with the RDA as QDF-R [6]. By the RDA, the covariance matrix is interpolated with an identity matrix by

$$\hat{\Sigma}_i = (1 - \gamma)\Sigma_i + \frac{\gamma}{d}\text{tr}(\Sigma_i)I, \tag{7}$$

where $0 < \gamma < 1$.

3 Orthogonal Quadratic Discriminant Functions (OQDF)

The MQDF2 obtained a good performance on the handwritten recognition. However, the number of principle eigenvectors k is hard to determined in advance. On the other hand, the MQDF model still has many parameters to be estimated. To overcome the shortcomings, we would like to improve the QDF approach and propose an orthogonal quadratic discriminant functions (OQDF) in this paper.

3.1 Motivations

Like the QDF and MQDF, the OQDF also assumes that the underlying density function is Gaussian. Furthermore, the OQDF assumes the underlying density functions of each two classes have a uniform shape. By taking an orthogonal transformation: such as rotating, reflecting and translating on class ω_a , we can obtain another class ω_b .

Mathematically, we can define an equivalent relation between two classes

$$\omega_a \sim \omega_b, \tag{8}$$

if there exists an orthogonal matrix Γ_1 and a vector v , s.t., $\forall x \in R^d$

$$P(\Gamma_1 x + v|\omega_a) = P(x|\omega_b). \tag{9}$$

The reflexivity, symmetry and transferability are easily proved. This equivalent relation is called **Uniform Shape Distribution (USD)** in this paper.

Theorem 1. $\omega_a \sim \omega_b \Leftrightarrow \Lambda_a = \Lambda_b$.

3.2 OQDF

Λ_i in Eq. (2) is replaced with $\Lambda_i^{new} = \Lambda^{new} = \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_d, \}$. Then Eq. (3) is rewritten as

$$g_3(x, \omega_i) = \sum_{j=1}^d \frac{1}{\sigma_j} [\phi_{ij}^T(x - \mu_i)]^2 + \sum_{j=1}^d \log \sigma_j. \tag{10}$$

As $\sum_{j=1}^d \log \sigma_j$ is independent of class label, then Eq. (10) can be simplified as

$$g_4(x, \omega_i) = \sum_{j=1}^d \frac{1}{\sigma_j} [\phi_{ij}^T(x - \mu_i)]^2. \tag{11}$$

The covariance matrix of ω_i satisfies $\text{Rank}(\Sigma_i) \leq \min(n_i - 1, d)$. For efficient computation, let $k = \min(n^{\min} - 1, d)$, where $n^{\min} = \min_i(n_i)$, the Eq. (11) can be rewritten as

$$g_o(x, \omega_i) = \frac{1}{\sigma} \|x - \mu_i\|^2 + \sum_{j=1}^k \left(\frac{1}{\sigma_j} - \frac{1}{\sigma}\right) [\phi_{ij}^T(x - \mu_i)]^2. \tag{12}$$

Eq. (12) is the proposed **Orthogonal quadratic discriminant functions (OQDF)**.

We developed three models, which are listed as follows

$$\text{OQDF-M: } \sigma_j = \left(\frac{1}{N} \sum_{i=1}^C n_i \lambda_{ij}\right)^{\gamma_1}, \tag{13}$$

$$\text{OQDF-E: } \sigma_j = (1 + e^{-\gamma_2 j})^{1/\gamma_2}, \tag{14}$$

$$\text{OQDF-L: } \sigma_j = \frac{1}{1 + \gamma_3 j}, \tag{15}$$

where $0 < \gamma_1, \gamma_2, \gamma_3 < 1$. And $\sigma = 2\sigma_k$.

The minimum classification error (MCE) criterion can be adopted to optimize the parameters of OQDF as in [13]. In this paper, we didn't implement the MCE for training the OQDF, instead we just justify the basic OQDF.

3.3 Efficient Computation

In the face recognition problem, the samples often have high dimensional features. The covariance matrix Σ_i of class ω_i is with size of $d \times d$, where d is the feature dimension. If $d > n_i$, n_i is the number of samples in ω_i , then the K-L transform can be accelerated.

Let matrix $A_i = [x_1 - \mu_i, x_2 - \mu_i, \dots, x_{n_i} - \mu_i]$, then $\Sigma_i = A_i A_i^T$. Assume λ_{ij} and $\bar{\phi}_{ij}$ are eigenvalue and eigenvector of $A_i^T A_i$, then λ_{ij} and $\phi_{ij} = A_i \bar{\phi}_{ij}$ are eigenvalue and eigenvector of matrix Σ_i respectively, since

$$A_i^T A_i \bar{\phi}_{ij} = \lambda_{ij} \bar{\phi}_{ij} \Rightarrow A_i A_i^T (A_i \bar{\phi}_{ij}) = \lambda_{ij} (A_i \bar{\phi}_{ij}). \tag{16}$$

This strategy has been adopted in the famous approach, Eigenfaces [17], for face recognition. In this way, the computational complexity can be reduced from $O(d^3)$ to $O(n_i^3)$.

4 Experiments and Discussions

To justify the proposed OQDF for face recognition problems, we compare the OQDF with some well known classifiers and improvements of the QDF. These approaches are listed in table 1. The Libsvm toolbox for matlab is used to train the SVM classifiers.

Table 1. Classifiers for comparison

Classifier	Descriptions
NN	Nearest Neighbor classifier [15].
NC	Nearest Centroid classifier by Eq. (5).
SVM-L	SVM with linear kernel, $K(x, x') = x^T x'$, [16]
SVM-R	SVM with RBF kernel, $K(x, x') = e^{-\ x-x'\ ^2/2\sigma^2}$
QDF	the basic QDF
QDF-R	$\gamma = 0.01$ (Eq. (7)) for all experiments.
MQDF3	k is determined by cross validation [11].
MQDF2	k and σ_i are determined by cross validation [3].
OQDF-M	$\gamma_1 = 0.3$ for all experiments.
OQDF-E	$\gamma_2 = 0.2$ for all experiments.
OQDF-L	$\gamma_3 = 0.5$ for all experiments.

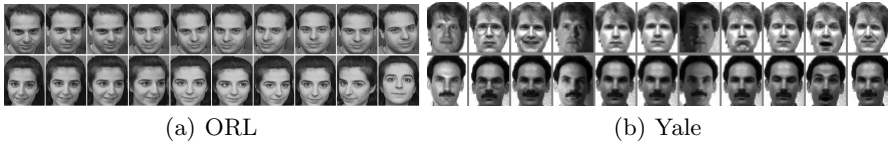


Fig. 1. Samples from ORL and Yale databases

4.1 Experimental Setup

In order to evaluate the proposed face recognition system, our experiments are conducted on two benchmark face databases: 1)The Olivetti Research Laboratory (ORL) database, 2) The Yale database. Fig. 1 shows some sample images from the ORL and Yale face database.

All the images were normalized to unit first before feature extraction. And the LST+FLD are employed to do feature extraction for all the compared classifiers on the two face databases.

4.2 Comparisons on ORL and Yale Database

Two small face databsets, ORL and Yale, are used to evaluate the classifiers. The ORL database contains face images from 40 individuals, each providing 10 different images. For some subjects, the images were taken at different times.

Table 2. Comparison on the ORL Database(mean±std-dev%). $d = 20$.

Classifier	G2/P8	G3/P7	G4/G6	G5/P5	G6/P4	G7/P3	G8/P2
NN	20.0±2.9	9.6±1.9	5.8±1.8	3.3±1.3	2.7±1.6	2.1±1.1	1.1±1.1
NC	20.0±2.8	11.3±2.4	6.2±2.1	3.9±1.4	2.8±1.4	2.5±1.5	1.3±1.2
SVM-L	20.2±2.9	11.4±2.3	6.1±1.9	3.5±1.4	2.8±1.3	2.1±1.1	1.1±1.2
SVM-R	20.1±2.9	11.4±2.3	5.8±1.9	3.1±1.2	2.4±1.3	1.8±1.1	0.9±1.0
QDF	89.5±2.0	84.9±2.4	79.4±2.6	73.9±2.4	71.0±3.5	67.5±4.7	64.9±4.6
QDF-R	43.7±6.6	22.2±6.6	10.0±3.3	5.0±1.8	3.6±1.7	2.8±1.5	1.4±1.3
MQDF3	65.4±8.5	53.6±8.6	39.2±7.6	32.3±6.1	28.4±6.5	24.5±6.2	20.8±6.0
MQDF2	20.4±2.5	11.7±2.3	6.2±1.9	3.6±1.5	2.7±1.5	2.0±1.2	0.9±1.2
OQDF-M	20.0±2.8	9.3±2.0	5.3±1.8	3.2±1.5	2.3±1.3	1.4±1.0	0.7±1.1
OQDF-E	19.9±2.8	9.3±1.9	5.3±1.5	3.2±1.5	2.4±1.5	1.4±1.0	0.8±1.0
OQDF-L	19.8±2.8	9.2±1.9	5.3±1.6	3.1±1.4	2.3±1.5	1.4±1.0	0.7±0.9

Table 3. Comparison on the Yale Database(mean±std-dev%). $d = 14$.

Classifier	G2/P9	G3/P8	G4/G7	G5/P6	G6/P5	G7/P4	G8/P3
NN	16.0±3.6	9.0±2.6	5.6±2.2	4.1±2.1	2.7±1.6	3.0±2.0	2.0±2.2
NC	13.8±3.5	8.4±2.5	5.2±2.2	4.0±1.7	2.5±1.7	3.1±1.8	2.1±2.0
SVM-L	12.4±3.0	8.8±2.4	5.5±2.2	4.0±2.1	2.7±1.6	3.0±1.9	2.3±2.3
SVM-R	14.9±3.2	8.8±2.4	5.6±2.2	4.0±2.0	2.7±1.6	2.9±1.9	2.2±1.9
QDF	63.2±9.4	22.2±7.0	21.3±12.9	22.0±17.1	20.0±24.3	23.7±23.4	35.6±32.8
QDF-R	17.2±3.4	10.7±2.9	11.1±17.2	6.8±2.1	6.0±2.7	6.5±2.8	3.6±3.3
MQDF3	29.0±8.2	19.4±5.6	14.3±6.3	12.2±4.3	12.7±5.4	12.7±5.7	15.2±8.8
MQDF2	16.7±3.5	9.8±2.6	10.8±17.2	6.8±2.3	6.0±3.1	7.2±3.4	6.8±4.1
OQDF-M	13.8±3.6	8.4±2.4	5.2±2.3	3.7±1.9	2.5±1.4	2.9±1.8	1.9±2.0
OQDF-E	13.6±3.5	8.3±2.3	5.1±2.2	3.7±1.8	2.5±1.4	3.0±1.8	1.8±1.9
OQDF-L	13.8±3.6	8.3±2.3	5.3±2.2	4.0±2.1	2.7±1.5	3.2±2.0	2.0±2.2

The images were taken with a tolerance for some tilting and rotation of the face of up to 20 degrees. Moreover, there is also some variation in the scale of up to about 10 percent.

Yale database contains 165 face images from 15 individuals, each providing 11 different images with different facial expression or configuration: center-light, w/glasses, happy, left-light, w/no glasses, normal, right-light, sad, sleepy, surprised, and wink.

The LST+FLD dimensionality reduction approach is employed to extract efficient features before training. For each Gp/Pq, we average the results over 50 random splits and report the mean as well as the standard deviation. Tables 2 and 3 show the results of different numbers of training samples on ORL and Yale databases. The experimental results show that the three OQDF classifiers have much lower error rates than the traditional QDF, MQDF approaches and nearest neighbor (NN), nearest centroid (NC). The OQDF classifiers are even superior to the SVM classifiers in most cases.

5 Conclusions

We proposed a new quadratic discriminant function, the orthogonal quadratic discriminant functions (OQDF), for face recognition. The covariance of each class ω_i is constraint to be $\Sigma_i = \Gamma_i \Lambda \Gamma_i^T$, where Γ_i is an orthogonal matrix and Λ is a diagonal matrix independent of class. Compared to the traditional QDF and MQDF models, the OQDF has much fewer parameters to be estimated. All experiments show that the three OQDF approaches have much lower error rates than the MQDF2 and the QDF-R model. The OQDF even outperforms the NN, NC and SVM classifiers on the two databases. The assumption of the OQDF can also be regarded as a regularizer for the classification models to solve the SSS problem.

However, the OQDF-E and OQDF-L models are not optimal models. The optimal σ_j can be found by some training algorithms, such as the MCE rule as in [4] (refer to Appedix).

Acknowledgment. This work is in part supported by the National High Technology Research and Development Program of China (863 Program), with grant number 2007AA01Z453, and partially supported by National Natural Science Foundation of China (NSFC), under grant number 60673020 and 60875080.

References

1. Er, M.J., Chen, W., Wu, S.: High-Speed Face Recognition Based on Discrete Cosine Transform and RBF Neural Networks. *IEEE Trans. Neural Networks* 16(3) (2005)
2. Heisele, B., Ho, P., Poggio, T.: Face Recognition with Support Vector Machines: Global Versus Component-Based Approach. In: *ICCV* (2001)
3. Kimura, F., Wakabayashi, T., Tsuruoka, S., Miyake, Y.: Modified Quadratic Discriminant Functions and its Application to Chinese Character Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 149–153 (1987)
4. Liu, C.L., Sako, H., Fujisawa, H.: Discriminative Learning Quadratic Discriminant Function for Handwriting Recognition. *IEEE Trans. Neural Networks* (2004)
5. Friedman, J.H.: Regularized Discriminant Analysis. *J. Am. Statist. Ass.* 84, 165–175 (1989)
6. Lu, J., Plataniotis, K., Venetsanopoulos, A.: Regularized Discriminant Analysis for the Small Sample Size Problem in Face Recognition. *Pattern Recognition Letters*, 3079–3087 (2003)
7. Wang, J., Plataniotis, K., Lu, J., Venetsanopoulos, A.: Kernel Quadratic Discriminant Analysis for Small Sample Size Problem. *Pattern Recognition*, 1528–1538 (2008)
8. Yu, H., Yang, J.: A Direct lda Algorithm for High-Dimensional Data with Application to Face Recognition. *Pattern Recognit.* 34, 2067–2070 (2001)
9. Gu, S., Tan, Y., He, X.: Laplacian Smoothing Transform for Face Recognition. *TPAMI* (submitted, 2008)
10. Duda, R., Hart, P.: *Pattern Classification*, 2nd edn. Wiley, New York (2001)
11. Moghaddam, B., Pentland, A.: Probabilistic Visual Learning for Object Representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 696–710 (1997)
12. Long, T., Jin, L.: Building Compact mqdf Classifier for Large Character Set Recognition by Subspace Distribution Sharing. *Pattern Recognition* 41, 2916–2925 (2008)

13. Juang, B.H., Katagiri, S.: Discriminative Learning for Minimum Error Classification. *IEEE Trans. Signal Processing* 40, 3043–3054 (1992)
14. Turk, M., Pentland, A.: Eigenfaces for Recognition. *J. Cognitive Neuroscience* 3, 71–86 (1991)
15. Cover, T., Hart, P.: Nearest Neighbor Pattern Classification. *IEEE Trans. Info. Theo.*, 21–27 (1967)
16. Fan, R.E., Chen, P.H., Lin, C.J.: Working Set Selection Using Second Order Information for Training Support Vector Machines. *Journal of Machine Learning Research* 6, 1889–1918 (2005)
17. Turk, M., Pentland, A.: Eigenfaces for recognition. *J. Cognitive Neuroscience* 3(1), 71–86 (1991)

Appendix: Discriminative Learning (DL) OQDF

Φ_i and μ_i can be trained also, however, due to the SSS problem, we only train σ_j .

Suppose $\{x_n, y_n\}_{n=1}^N$ are N training samples, $w = [w_1, w_2, \dots, w_{k+1}]^T$, with $w_j = \frac{1}{\sigma_j}$, $j = 1, 2, \dots, k$ and $w_{k+1} = \frac{1}{\sigma}$. Let $p_n^i = [p_{n1}^i, p_{n2}^i, \dots, p_{n(k+1)}^i]^T$, where

$$p_{nj}^i = [\phi_{ij}^T(x - \mu_i)]^2, j = 1, 2, \dots, k, \quad (17)$$

and

$$p_{n(k+1)}^i = \|x - \mu_i\|^2 - \sum_{j=1}^k [\phi_{ij}^T(x - \mu_i)]^2. \quad (18)$$

The OQDF decision function, as given in Eq. (12), can be rewritten as

$$g(x_n, \omega_i, w) = \sum_{j=1}^{k+1} w_j p_{nj}^i = w^T p_n^i. \quad (19)$$

We aim to find an optimal w , s.t.

$$g(x_n, \omega_c, w) < g(x_n, \omega_i, w), \text{ if } x_n \in \omega_c. \quad (20)$$

The misclassification measure of a pattern x_n from class ω_c is given by

$$h(x_n, w) = g(x_n, \omega_c, w) - \bar{g}(x_n, \omega_c, w) \quad (21)$$

where

$$\bar{g}(x_n, \omega_c, w) = \left[\frac{1}{C-1} \sum_{i \neq c} (w^T p_n^i)^{-\eta} \right]^{-\frac{1}{\eta}}, (\eta > 0). \quad (22)$$

When approaches η infinity, the misclassification measure becomes

$$h(x_n, w) = g(x_n, \omega_c, w) - g(x_n, \omega_r, w), \quad (23)$$

where is the discriminant function of the closest rival class:

$$g(x_n, \omega_r, w) = \min_{i \neq c} g(x_n, \omega_i, w) \quad (24)$$

The simplification of misclassification measure by setting $\eta \rightarrow \infty$ is helpful to speed up the learning process by stochastic gradient descent, where only the parameters involved in the loss function are updated on a training pattern. Embedding $h(x_n, w)$ into a sigmoid function, we get a continuous loss function $e(x_n, w)$ with respect to w ,

$$e(x_n, w) = \frac{1}{1 + e^{-\alpha h(x_n, w)}}, (\alpha > 0). \tag{25}$$

The empirical loss on the whole training set X is the summarization of the individual loss

$$E(X, w) = \sum_{n=1}^{\mathcal{N}} e(x_n, w). \tag{26}$$

$E(X, w)$ is then minimized using a generalized probability descent (GPD) algorithm. The OQDF-E model is taken as the initial w , and gradually better estimated is obtain by an iterative training scheme

$$w_{t+1} = w_t - \varepsilon_t \nabla E(X, w_t). \tag{27}$$

According to derivation rules,

$$\frac{\partial E(X, w)}{\partial w} = \sum_{n=1}^{\mathcal{N}} \frac{\partial e(x_n, w)}{\partial h(x_n, w)} \cdot \frac{\partial h(x_n, w)}{\partial w}, \tag{28}$$

where

$$\frac{\partial e(x_n, w)}{\partial h(x_n, w)} = \alpha e(x_n, w)(1 - e(x_n, w)), \tag{29}$$

$$\frac{\partial h(x_n, w)}{\partial w} = p_n^c - p_n^r. \tag{30}$$

LISA: Image Compression Scheme Based on an Asymmetric Hierarchical Self-Organizing Map

Cheng-Fa Tsai and Yu-Jiun Lin

Department of Management Information Systems
National Pingtung University of Science and Technology, Pingtung, Taiwan, 91201
{cftsai, m9656013}@mail.npust.edu.tw

Abstract. A Kohonen network, also called Self-Organizing Map (SOM), is a competitive learning network, and is appropriate for solving an image compression problem owing to its ability to generate high-quality compressed images. However, SOM has a large computation cost, making it impractical due to a lengthy training process. Hence, the Hierarchical Self-Organizing Map (HSOM) had been presented and found to reduce computation cost. Although a hierarchical architecture speeds up SOM, HSOM is still not practical enough because of a high compression cost. Therefore, this investigation employs a hybrid scheme to increase the efficiency and effectiveness of HSOM. Simulation results reveal that the proposed algorithm is much more efficient and effective than other algorithms, such as LBG, SOM, and HSOM.

Keywords: image compression, vector quantization, SOM, HSOM.

1 Introduction

Internet use has grown exponentially in recent years. However, restrictions of network bandwidth and storage space limit the size of files transmitted across the Internet. In summary, the file size must remain small in order to maintain a fast response after “click”, making image compression into a widely researched topic. Image compression approaches are classified as lossy and lossless. The conventionally utilized lossy method is Vector Quantization (VQ) [1], [2], since VQ approaches generate compressed images with high compression ratio and high image quality. For instance, the color level of a gray uncompressed image is 8bpp (bits/pixel). By contrast, a compressed image with a size of 512×512 generated by VQ approaches with vector of 4×4 and a codebook size of 1024, has a low bit rate is about 0.625 bpp. The small index-book is valuable for fast transmission in the cyberspace. Correlative researchers have developed various VQ approaches, including LBG [3], SOM [4], [5], [6], [7] and HSOM [8]. The most popular VQ scheme is Self-Organizing Map (SOM), since it discovers better codebooks than other VQ approaches. However, SOM approaches have high computation cost, making them impractical despite obtaining high-quality compressed images. Therefore, this investigation proposes a hybrid scheme involving LBG and an Improved SOM based on an Asymmetric hierarchical architecture.

The proposed approach is called “**LISA**”. The rest of this paper is organized as follows. Section 2 introduces the above VQ algorithms. Section 3 then describes the proposed method in detail, including its concept and step-by-step procedure. Next, Section 4 summarizes the simulation results of *PSNR* and *execution time*. Conclusions are finally drawn in Section 5.

2 Related Works

This section presents the essential concepts of Vector Quantization; describes several well-known VQ techniques, namely LBG, SOM and HSOM, and briefly summarizes their strengths and weaknesses. Finally, the common definitions of image quality measures are shown.

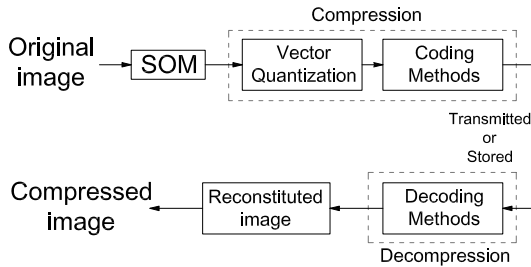


Fig. 1. The image compression process in the neural network system

Fig.1. illustrates the image compression process in the neural network system. **Vector Quantization**, which is adopted mainly to design codebooks, is a widely-employed lossy compression method that can decrease the compression rate and preserve good quality following compression. Fig. 2 depicts the coding process of VQ. An $N \times N$ gray image is first divided into $n \times n$ blocks, forming a block set V , which is converted into code-vectors, as $V = v_1, v_2, \dots, v_{n \times n}$. A codebook M is composed of k code-words, $M = CB_1, CB_2, \dots, CB_k$, where CB_k indicates the k th code-vector in the codebook. A code-vector represents an index-value in the index-book. When transmitting images in the network, only the codebook and the index-book need to be transmitted, rather than entire pixels of original image, thus lowering the storage space and transmission time.

The **LBG** algorithm, also called K-means, was first developed by Linde Y., Buzo A. and Gray R.M. in 1980, and has two major steps, namely data grouping and updating centroid. In summary, LBG assigns code-vectors in the codebook by continuously comparing the distance between the training dataset and centroid of the group until the variation of average distortion is below the stopping threshold. The simplicity of the LBG algorithm means that it can obtain the codebook efficiently. However, choosing an initial codebook randomly may cause the algorithm to fall into a local optimum, possibly making the final result unstable.

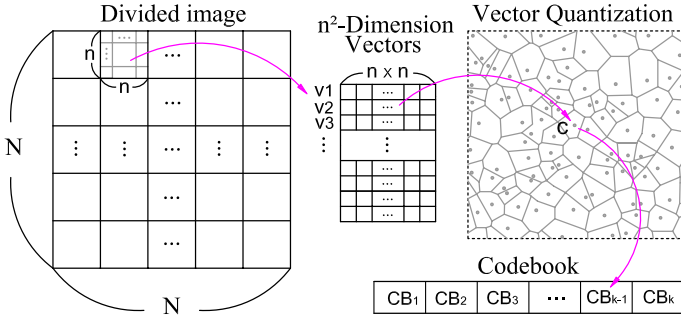


Fig. 2. The coding process of VQ

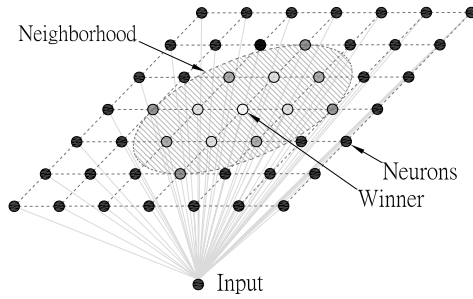


Fig. 3. The operation of “neighborhood” in SOM

Kohonen proposed a **Self-Organizing Map (SOM)** for unsupervised neural networks in 1980. Although SOM is a competitive learning network, it is not based on “winner takes all”. SOM utilizes the concept of “neighborhood”, in which neurons neighboring the winning neuron are also activated. Fig.3. displays the concept of the “neighborhood” of SOM. The neighborhood is typically obtained by a Gaussian function. Overall, SOM usually produces good results, but has to perform a full search on the training data set, and therefore incurs a fairly high computation cost.

Hierarchical Self-Organizing Map (HSOM) was proposed by Barbalho in 2001. Importantly, HSOM can decrease the time complexity of SOM from $O(n)$ to $O(\log n)$. However, HSOM still has the limitation that each sub map has the same size, making it liable to fall into local optima. Furthermore, the data distribution of HSOM cannot be represented effectively. Some objective measures of verifying the compressed image quality are adopted. For instance, mean square error (MSE) and Peak Signal-to-Noise Ratio (PSNR) are commonly used, and are formulated as follows.

$$PSNR = 10 \times \log_{10}\left(\frac{255^2}{MSE}\right) \tag{1}$$

$$MSE = \frac{1}{M \times N} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [\hat{I}(x, y) - I(x, y)]^2 \tag{2}$$

3 The Proposed Algorithm

The standard Hierarchical SOM is framed in a symmetric architecture of two levels with a 1-D string neural network. Therefore, the neuronal number in the first level is equal to that in each sub-group in the second level. Fig. 4 depicts this concept of symmetric two levels HSOM, and $N = M$ in traditional HSOM. The proposed algorithm performs three major steps to accelerate HSOM and enhance quality. The first major step, to speed up the entire training process, is described as follows. An asymmetric structure ($N > M$) is adopted. All data vectors are split into many sub-groups in level 1, leaving only few data vectors in each sub-group of level 2. This approach speeds up the training process of level 2 successfully, but the large number of neurons in the first level causes a lengthy training process in level 1. To eradicate this problem, the proposed algorithm utilizes only few data vectors that generated by LBG algorithm with stopping threshold $\varepsilon=0.01$ to train neural map, thus significantly reducing the computational cost of the training process in level 1. The next section lists the parameters of numbers of neurons at level 1, and of important data vectors be selected.

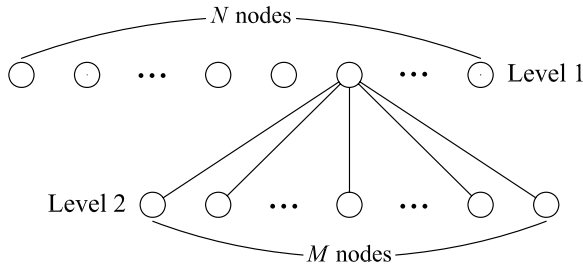


Fig. 4. The architecture of HSOM

The second major step is to employ a modified SOM algorithm rather than conventional SOM to accelerate the training process of each group. The number of training epochs for conventional SOM needs to be set in advance, and is fixed for each group. In summary, each group has various distributions. Notably, there are $N + 1$ groups in the two levels hierarchical architecture. Therefore, training iterations of each group could be expected to be different. To overcome this problem, the proposed method calculates the average distortion after completing each training iteration, and employs Eqn. (3) to derive the stopping threshold ε representing the variation of average distortion. The average distortion is determined by Eqn. (4), where x_num denotes the number of samples. Finally, the training is completed if ε is less than a pre-setting value that would

be discussed in next section. This stopping mechanism is similar to that of the LBG algorithm, except our proposed scheme has a limit of at most 500 epochs. Notably, it is necessary to avoid trapping an infinite loop.

$$\varepsilon = \left| \frac{\bar{D}(t) - \bar{D}(t+1)}{\bar{D}(t)} \right| \quad (3)$$

$$\bar{D}(t) = \frac{\sum d(x, \text{centroid}(x))}{x_num} \quad (4)$$

The third major step solves the problem that the fixed neuronal number in each sub map may cause the algorithm to fall into a local optimum. A dynamic scheme that assigns neuronal number within the proportion of squared Euclidean distance of each group in second level, and with a limit of at least 1, is utilized. Eqn. (5) represents the squared Euclidean distance, where i denotes the vector number; r indicates the dimension of vector, and all data vectors are in $\mathbb{R}^{4 \times 4}$ herein; c is the centroid of the cluster. Eqn. (6) computes the number of neurons (N_g) in each sub-group, where g devotes g th sub-group.

$$d = \sum_{i=1}^n \sum_{r=1}^{4 \times 4} (x_{ir} - c_r)^2 \quad (5)$$

$$N_g = \left\lceil \frac{d_g}{\sum d_g} \times \text{codebook_size} \right\rceil \quad (6)$$

The procedure of LISA algorithm can be described step by step below:

1. Initialize all parameters including ε : Stopping threshold, N : neuronal number of first level, k : number of cluster and η_0 : initial learning rate.
2. Estimate k virtual samples (y_k) using LBG ($\varepsilon=0.01$). Due to the limit on paper length, this work does not discuss the methodology of LBG in detail. Please refer to paper [3].
3. Utilize virtual samples (y_k) and modified SOM with a stopping mechanism to train a randomly established neural network of size N . The training process is listed below with *search winner* and *weight adaptation*.

Search the winner (w_j^*) by comparing the distance between virtual samples (y_k) and neurons, as in Eqn. (7), where $\|\cdot\|$ represents the Euclidean distance and $w_j(t)$ depicts the weight of j th neuron at training times t .

$$w_j^* = \arg \min \|y_k - w_j(t)\|, \quad j = 1, 2, \dots, N \quad (7)$$

Weight adaptation: The weights of the neurons neighboring winner are all adapted by Eqn. (8).

$$w_j(t+1) = w_j(t) + \eta(t) \times h_{ji}(t) \times [y_k - w_j(t)] \quad (8)$$

$$h_{ji}(t) = \exp\left(-\frac{r}{R}\right) \quad (9)$$

Eqn. (9) derives the neighborhood function, where $r = \|j - j^*\|$ and R denotes the radius of neighborhood.

4. Update parameters R and η by linear decreasing, as in conventional SOM.
5. Repeat Steps 3-5 until the stopping criteria are reached.
6. Split all data vectors into many sub-groups according to neural map of level 1 by comparing distance between data vector and each neuron, and assigning data vector to the group (i.e. which the neuron with minimum distance), until all data vectors are assigned to each group. Calculate the number of neurons in each group by Eqn. (5) and Eqn. (6).
7. Choose one sub-group, while initializing the neural map randomly with its own size. Train the neural map by following step 8.
8. Utilize all true samples (x_i) belonging to the selected sub-group, and modify SOM to train the neural network. The training process has two step of *searching winner* and *weight adaptation* are listed below.
Search winner (w_j^*) by comparing distance between true samples (x_i) and neurons, as in Eqn. (10), where $i \in$ selected sub-group and N_g denotes the neuronal number of g th group.

$$w_j^* = \arg \min \|x_i - w_j(t)\|, \quad j = 1, 2, \dots, N_g \quad (10)$$

Weight adaptation: The weights of the neurons neighboring winner are determined by Eqn. (11).

$$w_j(t+1) = w_j(t) + \eta(t) \times h_{ji}(t) \times [x_i - w_j(t)] \quad (11)$$

9. Linearly decrease parameters R and η .
10. Repeat Steps 8-9 until the stopping criteria are satisfied.
11. Repeat Steps 7-10 until all sub-groups are trained.

4 Experiment and Analysis

The experiment comprising quality of compressed images and time cost of the presented LISA algorithm were demonstrated. The program of each algorithm was conducted in a Java-based program and ran on a desktop computer with 2GB RAM and an Intel T7300 2.0 GHz CPU on Microsoft Windows XP professional operational system. Six gray images with image size of 512×512 were employed involving Lena, Airplane, Boat, Peppers, Ann and Sweets. Simulation results were calculated with the average of 30 rounds. For fair comparison, the parameters of HSOM approach were set as in paper [8]. Moreover, the stopping threshold (ε) of the LBG was set to 0.0001, while the training epoch was set to 200 for 1D SOM. For the proposed LISA, the number of clusters was set to 256 among every case, while the stopping thresholds (ε) in the first and second levels were set to 0.00001 and 0.000001, respectively. The learning rate in all cases was set to 1, and the numbers of neurons in the first level of codebooks with size 128, 256, 512 and 1024 were set to 40, 60, 120 and 240 respectively.

The test codebook sizes were 128, 256, 512, and 1024. All test images were grayscale. Table 1 summarizes the simulation results (PSNR and time cost) for LISA, LBG, SOM and HSOM using six images with codebook sizes of 128,

Table 1. The simulation results (PSNR and Time Cost) for LISA, LBG, SOM and HSOM. Boldface depicts the best one, while N/A denotes not-available.

Image	Codebook Size	PSNR (in dB)				Time Cost (in second)			
		LISA	LBG	ID-SOM	HSOM	LISA	LBG	ID-SOM	HSOM
Lena	128	29.579	29.569	29.686	N/A	9.642	8.394	142.09	N/A
	256	30.671	30.468	30.589	30.636	12.085	16.693	283.42	56.080
	512	31.816	31.272	31.477	N/A	13.643	28.215	563.08	N/A
	1024	33.235	32.106	32.436	32.973	17.082	45.065	1118.3	113.79
Airplane	128	29.239	28.839	29.320	N/A	8.103	12.891	142.38	N/A
	256	30.284	29.615	30.211	30.224	10.476	20.234	281.88	56.336
	512	31.343	30.458	31.133	N/A	11.855	27.254	561.63	N/A
	1024	32.563	31.452	32.166	32.472	15.916	38.515	1124.9	114.91
Boat	128	29.157	29.132	29.345	N/A	8.556	14.178	141.94	N/A
	256	30.206	29.935	30.247	30.166	10.577	21.082	282.02	55.771
	512	31.304	30.754	31.222	N/A	12.400	30.458	562.50	N/A
	1024	32.518	31.643	32.284	32.455	16.456	42.124	1116.4	114.08
Peppers	128	29.701	29.674	29.787	N/A	9.489	9.019	142.58	N/A
	256	30.660	30.488	30.607	30.627	11.762	16.013	283.02	55.847
	512	31.620	31.223	31.396	N/A	13.708	26.168	564.28	N/A
	1024	32.672	31.985	32.308	32.573	17.975	38.181	1117.1	114.17
Ann	128	28.172	28.213	28.332	N/A	8.957	7.858	140.67	N/A
	256	29.249	29.183	29.339	29.254	11.423	14.782	284.75	55.674
	512	30.345	30.120	30.279	N/A	12.904	23.101	562.92	N/A
	1024	31.535	31.107	31.384	31.526	16.600	35.407	1115.9	113.84
Sweets	128	29.684	29.630	29.834	N/A	9.046	7.198	141.28	N/A
	256	30.847	30.641	30.853	30.822	11.263	12.171	281.53	56.072
	512	32.088	31.625	31.888	N/A	12.911	23.140	559.14	N/A
	1024	33.514	32.645	32.933	33.341	16.325	36.743	1117.6	114.29

256, 512, and 1024. Notably, the left-hand side of the table indicates the PSNR comparison, while the right-hand side of the table represents the time cost comparison. Furthermore, the notation of “N/A” in Table 1 denotes “not available”. The codebook setting in HSOM must be N^2 , where N indicates the square root of the codebook size. Therefore, the only possible codebook sizes are 256 and 1024, since N should be an integer number.

Figs. 5 and 6 display the broken line graph of comparison of PSNR and time cost for LISA, LBG, SOM and HSOM using six gray images with test codebook sizes of 128, 256, 512, and 1024. These results reveal that the proposed LISA generated compressed image with best quality and had the lowest computation cost at codebook size ≥ 512 . Although the image quality of LISA in codebook size=128 was slightly lower than SOM, it still found to be faster significantly than SOM. Moreover, the PSNR value of compressed images generated by each algorithm with codebook size 128 was fairly low. Therefore, the quality of images

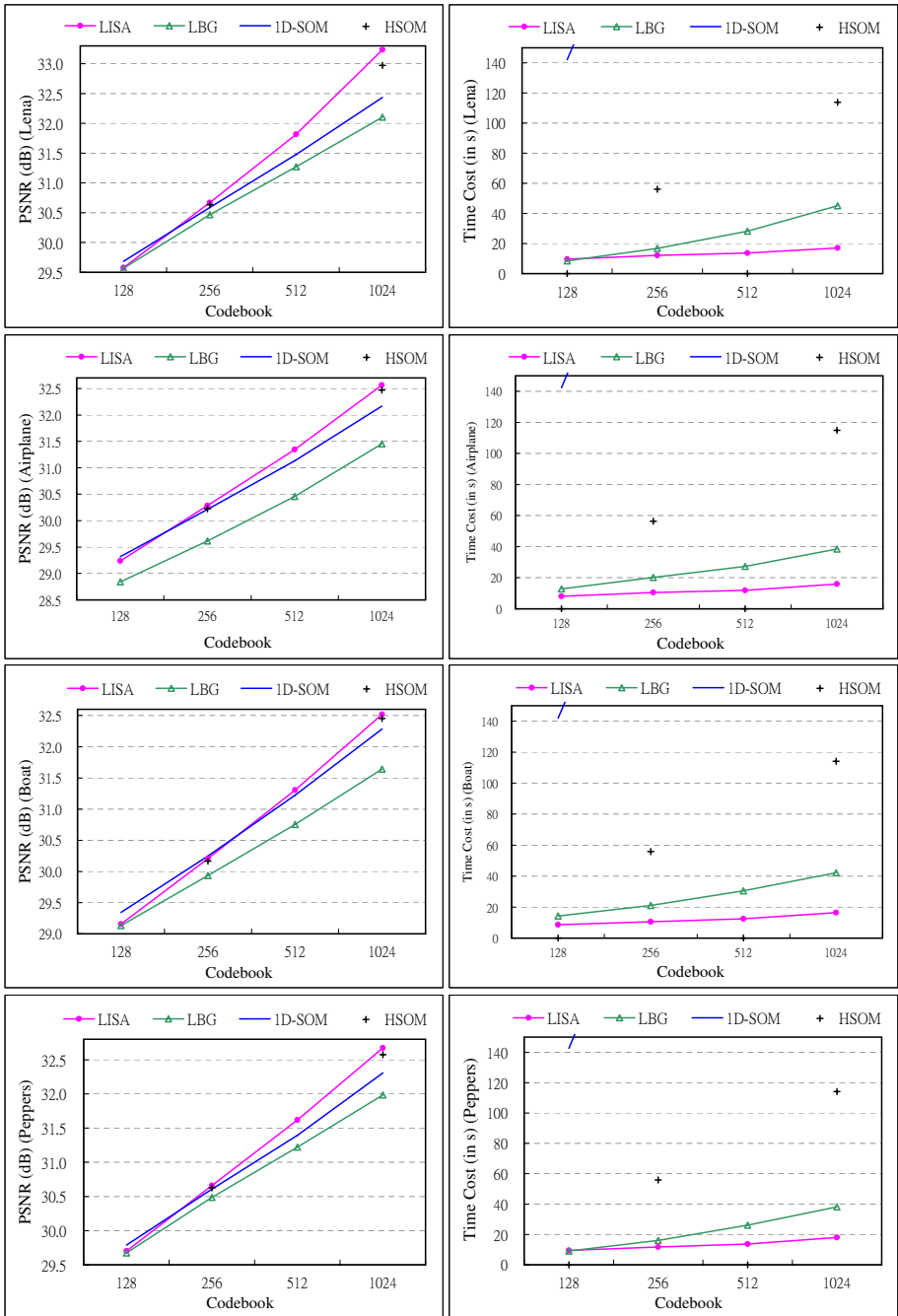


Fig. 5. The broken-line graph of PSNR (in dB) and time cost (in second) for LISA, LBG, SOM and HSOM using two gray Lena and Airplane images with 128, 256, 512, and 1024 test codebook sizes

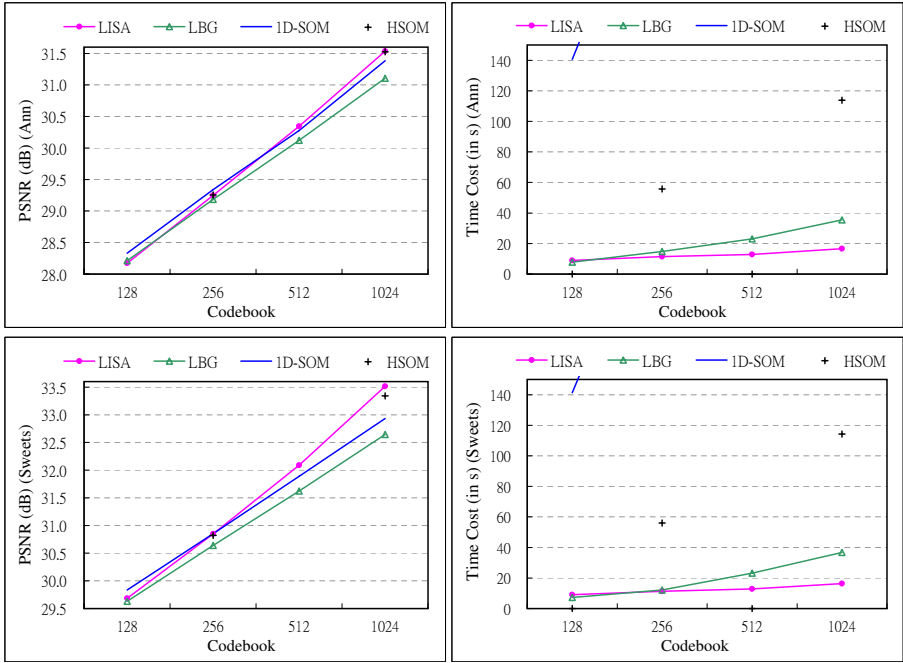


Fig. 6. The broken-line graph of PSNR (in dB) and time cost (in second) for LISA, LBG, SOM and HSOM using two gray Ann and Sweets images

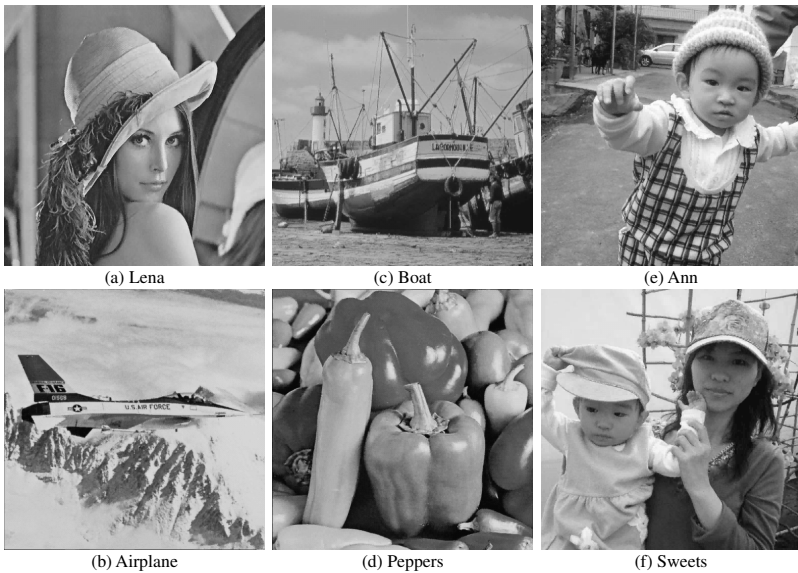


Fig. 7. The compressed images of Lena, Airplane, Boat, Peppers, Ann and Sweets generated by LISA with codebook size of 1024 and vector of 4×4

with codebook size of 128 would not be acceptable to end users, owing to excessive distortion. Fig.7. shows the compressed images of Lena, Boat, Airplane, Peppers, Ann and Sweets generated by LISA with codebook size 1024 and vector 4×4 . These images demonstrate that the compressed and original images look very similar to human eyes. In sum, these images have high quality and low storage space, making them likely to be acceptable to most end users. For limitation of paper length, there were only several figures and tables to demonstrate the performance of the proposed algorithm.

5 Conclusion

Simulation results indicate that LISA is a faster algorithm than the tested LBG, SOM and HSOM approaches, since it utilizes a modified SOM with asymmetric hierarchical structure and dynamic neuron number assignment to train the neural network. Moreover, LISA has better compressed image quality than the other tested algorithms, owing to the design of flexible stopping mechanism and dynamic neuron number assignment policy. In summary, the proposed LISA algorithm generates compressed images efficiently and effectively.

Acknowledgement. The author would like to thank the National Science Council of Republic of China, Taiwan for financially supporting this research under contract no. NSC 96-2221-E-020-027.

References

1. Gray, R.M.: Vector Quantization. IEEE ASSP 1(2), 4–29 (1984)
2. Sayood, K.: Introduction to Data Compression, 2nd edn. Morgan Kaufmann, San Francisco (2000)
3. Linde, Y., Buzo, A., Gray, R.M.: An algorithm for vector quantization design. IEEE Trans. Commun. COM-28, 84–95 (1980)
4. Kohonen, T.: Self-organizing maps. Springer, Berlin (1995)
5. Kohonen, T.: Self-organizing map. Proceedings of the IEEE 78(9), 1464–1480 (1990)
6. Madeiro, F., Vilar, R.M., Neto, B.G.A.: A Self-Organizing Algorithm for Image Compression. In: Proceedings of Vth Brazilian Symposium on Neural Networks, pp. 146–150 (1998)
7. Kangas, J., Kohonen, T.: Developments and applications of the self-organizing map and related algorithms. Mathematics and Computers in Simulation 41, 3–12 (1996)
8. Barbalho, M., Duarte, A., Neto, D., Costa, A.F., Netto, L.A.: Hierarchical SOM applied to image compression. In: Proceedings of International Joint Conference on Neural Networks, pp. 442–447 (2001)

A Method of Human Skin Region Detection Based on PCNN

Lijuan Duan¹, Zhiqiang Lin¹, Jun Miao², Yuanhua Qiao³

¹ College of Computer Science and Technology, Beijing University of Technology,
Beijing 100124, China

² Key Lab of Intelligent Information Processing, Institute of Computing Technology,
Chinese Academy of Sciences, Beijing 100190, China

³ College of Applied Science, Beijing University of Technology,
Beijing 100124, China

ljduan@bjut.edu.cn, ant_123@emails.bjut.edu.cn, jmiao@ict.ac.cn,
qiaoyuanhua@bjut.edu.cn

Abstract. A method of human skin region detection based on PCNN is proposed in this paper. Firstly, the input origin image is translated from RGB color space to YIQ color space, and I channel image is obtained. Secondly, we use the synchronous pulse firing mechanism of pulse coupled neural network (PCNN) to simulate the skin region detection mechanism of human eyes. Skin and non-skin regions are fired in different time. Therefore, skin regions are detected. Our comparison with other methods shows that the proposed method produces more accurate segmentation results.

Keywords: Skin Region Detection, PCNN, YIQ Color Space.

1 Introduction

Human skin detection is very important for many applications, such as face, hands gesture and human body detection or recognition in computer vision. It is widely used in the fields of human and machine interactive interface, access control, video monitoring and Internet pornographic image filtering.. The human skin detection methods based on color are simple, fast and intuitional. On the other hand, they are not sensitive to changes of shape and angle of view. Many researchers have focused on it [1-3]. Angelopoulou[1] indicated that human skin color distribution was consistent in biological and physical aspects. In other words, although different races have different skin colors, the hue of human skin is mostly similar when the influence of luminance and the environment is considered, which means human skin colors can congregate in a small color space.. Zhang et al. [2] pointed out that, I channel in YIQ color space has a good clustering characteristics for the human skin color in spite of the difference of the human race, the age or the gender.. It was obtained that human skin colors located in I channel was from 20 to 90 by some statistic experiments [2]. These two methods have the low detection performance under various illumination conditions. Tao et al. [3] proved that the characteristics of human skin pixels in RGB color space is that R value is larger than B value, and

B value is larger than G value, which is stable for various races and illumination conditions. However this method doesn't consider the relationship between neighboring pixels in terms of dealing with every pixel separately.

In order to overcome the problems above mentioned and detect skin regions of different human race efficiently, a novel human skin detection method is proposed based on the clustering characteristics of human skin in YIQ color space and the synchronous pulse firing mechanism of pulse coupled neural network(PCNN).. It can be used to detect human skin area in complex backgrounds. Even though the high light or shadow imposed on the human skin area, this method can also work well.

The paper is organized as follows. Section 2 introduces PCNN model and color space. Section 3 describes the framework of method proposed in this paper. Section 4 shows the experiment results. A discussion is given in Section 5.

2 PCNN Model

In 1990, PCNN is proposed by Eckhorn [4], which explains the experimentally observed synchronous activity among neural assemblies in the cat cortex induced by feature dependent visual activity. PCNN has interesting output, which differs from neural network composed of rate-coding neuron, since PCNN neuron can code information toward time axis. Subsequently PCNN has been used into image processing such as segmentation and fusion [5]. Some researchers modified the linking field network, and then it became the pulse coupled neural network [6] [7].. Fig. 1 shows a basic PCNN neuron model.

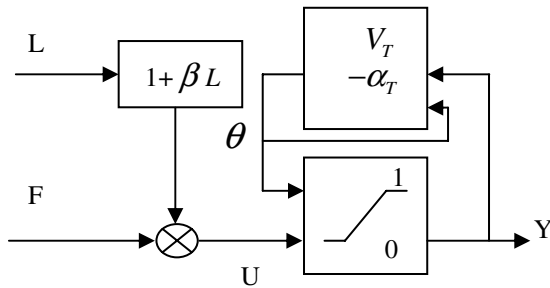


Fig. 1. A PCNN neuron model

The model has three main parts: the receptive fields, the modulation product, and the pulse generator [4]. It can be described by a group of equations [6].

$$F_i = \sum_j M_{ji} Y_j(t) \otimes \phi_{ji}(t) + I_i \dots \tag{1}$$

$$L_i = \sum_k W_{ki} Y_k(t) \otimes \phi_{ki}(t) + J_i \dots \tag{2}$$

$$U_i = F_i(1 + \beta_i L_i) . \quad (3)$$

$$Y_i(t) = \text{Step}(U_i - \Theta_i) .. \quad (4)$$

$$\Theta_i = -\alpha_T \Theta_i + V_T Y_i(t) . \quad (5)$$

The neuron receives input signals from other neurons and from external sources through the receptive fields. The signals include pulses, analog time-varying signals, constants, or any combination. Then the signals are divided into two channels. One is feeding channel, the other is linking channel.. In the modulation part the linking input is weighted with β_i and added a constant bias, then multiplied with the feeding input. The internal activity U_i is the output of the modulation part. In succession, the pulse generator compares U_i with a threshold θ_i . If U_i is larger than θ_i , the neuron will emit a pulse. It is also called ‘fire’. Otherwise, it will not fire. With reference to equation (6), Y_i is the output. At last, the pulse generator adjusts the threshold θ_i . If the neuron has fired, θ_i will be increased to a large value; otherwise, θ_i will decay (with reference to equation (5)).

Pulse output will be delivered to adjacent neurons. If adjacent neurons have similar intensity with neuron i , they will fire together because of pulse coupled action [5]. In this case, we call that neuron i captures the adjacent neurons. Finally the neuron i and the similar adjacent neurons will emit synchronous pulses. This is the theoretical foundation of PCNN for image segmentation.

Usually, when using PCNN to segment images, a single layer two-dimensional network is designed. In the network, the neurons and the pixels are in one to one correspondence. So, in this paper, one neuron is equal to a pixel.

3 Framework of Human Skin Region Detection Based on PCNN

The framework of human skin region detection based on PCNN is as Fig.2. Firstly, the input origin image is translated from RGB color space to YIQ color space, and I channel image is obtained. Secondly, we use PCNN to segment images. In order to decide the threshold in PCNN, the histogram of I channel is used to identify the range of I value. It is dynamic and adaptive I scope decision method, and it can segment image according to the image’s character, so that it is much objective. Finally, we can binary the result of PCNN multi-value segment.

3.1 Converting to YIQ Color Space and Getting I Channel Image

In YIQ color space, I channel can describe the change from orange to cyan, and Q channel can describe the change from purple to yellow-green. When we convert image from RGB color space to YIQ color space, we can divide the luminance information from hue information, then we can deal with images with light information and hue information separately. Zhang et al. [2] pointed out that, I channel in YIQ color

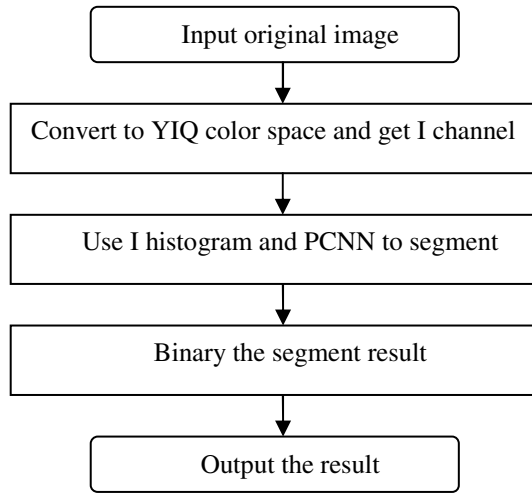


Fig. 2. Framework of human skin region detection based on PCNN

space has a good clustering characteristics for the human skin color in spite of the difference of the human race, the age or the gender.. It was obtained that human skin colors located in I channel was from 20 to 90 by some statistic experiments [2]. Therefore, we adopt YIQ color space in skin region detection. The conversion formulas are as follows:

$$Y = 0.299R + 0.587G + 0.114B \quad (6)$$

$$I = 0.596R - 0.275G - 0.321B \quad (7)$$

$$Q = 0.212R - 0.523G + 0.311B \quad (8)$$

3.2 Segmenting I Channel Image by Using PCNN

Usually, when using PCNN to segment images, a single layer two-dimensional network is designed. In the network, the neurons and the pixels are in one to one correspondence. So, in this paper, one neuron is equal to a pixel. Pulse output will be delivered to adjacent neurons. If adjacent neurons have similar intensity with fired neuron, they will fire too because of pulse coupled action. In other words, PCNN method thinks about relationship between neighboring pixels. It is coherent with human vision mechanism that the similar color should be segmented into an area block whether the conditions of illuminations are. Usually, the background and the target are much different, so the peaks of backgrounds and targets are different in histogram. Therefore, we can regard the lowest value between neighboring peaks in histogram as threshold in PCNN to filter some areas those are background or non-skin region obviously. It can also accelerate PCNN speed and improve the performance. Based on this method, the pixels in first obvious skin region will be

fired synchronously in advance. Then second obvious skin region fired subsequently. And so on, I channel image will be segmented into several regions.

3.3 Binary the Result of PCNN Segmentation

Because of the illumination condition and other reasons, skin blocks in one image will create several peaks in I histogram. A big skin region in original image will be separated into several small region based on the 3.2 section. However, I values of them are very close. So we can binary the result of PCNN multi-value segment.

The main idea is drawing the histogram of multi-valve segmented result image. Then, the trough between peaks in histogram is obtained. In order to represent non-skin pixels, the pixels whose values are smaller than the trough are regarded as background and labeled into zero, while the pixels whose values are larger than the trough are labeled into 1 and represented skin.

4 Experiments

In order to demonstrate the performance of the proposed method, some experiments are performed. We compare our method with that in paper [3], which segmented in YIQ color space and processed images with pixels. The main idea in reference 3 is as following. First, it converted multicolor images from RGB color space to YIQ color space and got I channel images. Then it checked every pixel's value in I channel. If I value of a pixel is between 20 and 90, it is labeled as skin pixel; otherwise it is a non-skin pixel. In order to display experimental results, the white pixels represent human skin; the black pixels represent non-human-skin.

Fig.3 shows the sample images from image library and internet in our experiments. Fig.3a is about yellow race people, and the background is similar to human skin color. Fig.3b is about white people. In Fig.3d, different areas in the picture are different illuminance condition. Fig.3c is got from Internet.

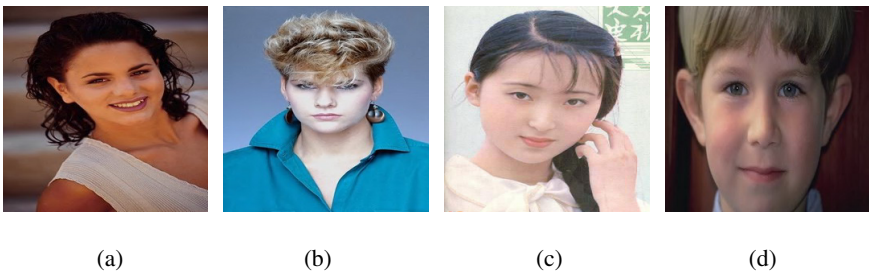


Fig. 3. Original images

Fig.4 shows the results by using the method mentioned in paper [3]. It can be found that some backgrounds of the picture are regard as human skin color. And the eyes are also regarded as human skin, as shown in Fig.4a. For the picture in Fig.4b, only a patch of the face area is segmented, while the neck and breast regions are not found. The main reason is that the model mentioned in paper [3] is not suitable for white race face detection. In Fig.4c, the mouth is regard as skin. In the last picture,

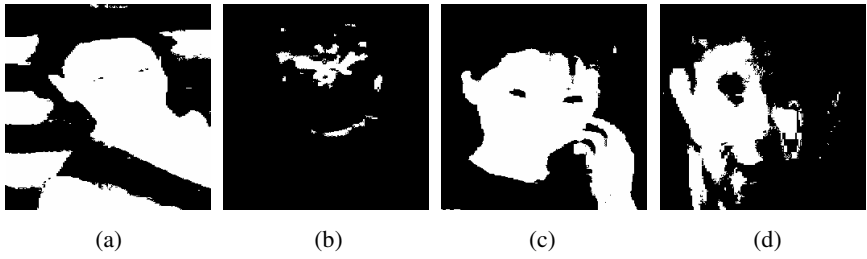


Fig. 4. The results of Using I Channel to do binary segmentation [3]

because there is shadow in the right face, it is much darker than the left face. As shown in Fig.4d, the right face cannot be detected by the method in paper [3].

Fig.5 shows the histogram of different images' I channel, we can see usually there are two obvious peaks, and use the trough as a threshold in PCNN segmentation.

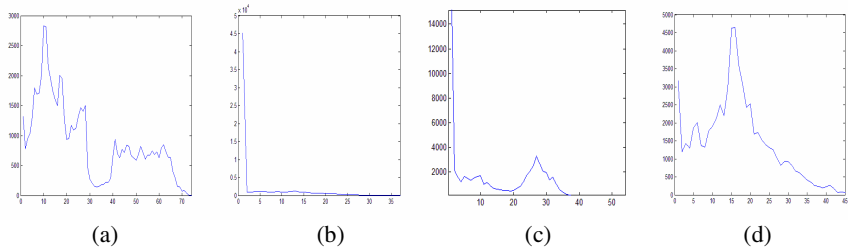


Fig. 5. I Channel histogram

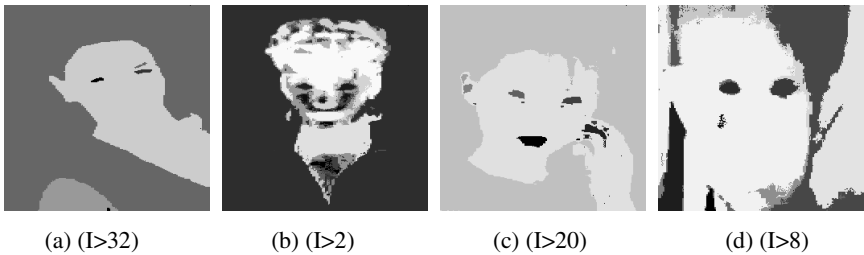


Fig. 6. The result of PCNN multiple value segmentation

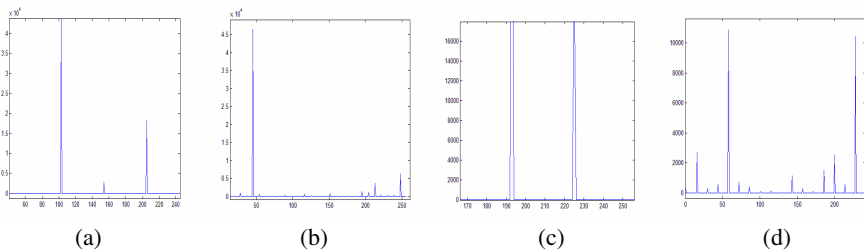


Fig. 7. Histogram of PCNN segmentation

Fig.6 shows the results of PCNN segmentation method mentioned in section 3.2. Fig.7 is the histograms of PCNN segmentation result, and Fig.8 is the last result.

As shown in Fig.8a, we distinguish the skin regions with background, the eyes and even eyebrow of the people of the image in Fig.3a. For the image in Fig.3b, by using PCNN we do multiple value segmentation, and get several patches of skin area. After binary the result of PCNN segmentation, it emphasizes the non-skin area as shown Fig.8b, and gets a better result. But the problem is that it cannot distinguish the brown hairs from skin; this is another problem we need to solve. For the picture in Fig.3c, we can distinguish the mouth from face as shown in Fig.8c. As to see the original image in Fig.3d, we can see that it has a complex illumination condition. Our method can detect the skin area easily, as shown in Fig.8d. It is obviously better than the result mentioned in Fig.4d.

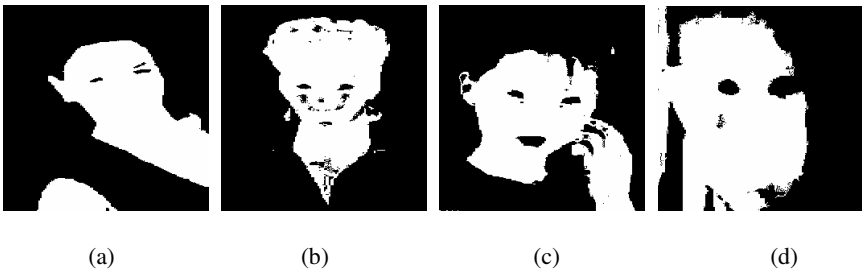


Fig. 8. Binary result based on PCNN multiple value segmentation

5 Discussion

Inspired by the synchronous pulse firing mechanism, we proposed a new method to detect human skin region in this article. We use the pulse coupled neural network (PCNN) on I channel image to segment skin and non-skin region.. Experiments show that this method can detect most skin areas in the images in spite of high illumination, shadow or people races. The current method is failed to distinguish the brown hair region from skin regions. The further work is to combine texture features to detect human skin region and remove the influence of other elements, such as brown hair.

Acknowledgements. This research is partially sponsored by Natural Science Foundation of China (Nos.60673091 and 60702031), Hi-Tech Research and Development Program of China (No.2006AA01Z122), Natural Science Foundation of Beijing (No.4072023), Beijing Municipal Education Committee (No.KM200610005012), Beijing Municipal Foundation for Excellent Talents (No. 20061D0501500211) and National Basic Research Program of China (No.2007CB311100).

References

1. Angelopoulou, E.: Understanding the Color of Human Skin. In: Proceedings of the SPIE Conference on Human vision and Electronic Imaging VI. SPIE, vol. 4229, pp. 243–251 (2001)
2. Zhang, H., Zhao, D., Gao, W., Chen, X.: Combining Skin Color Model and Neural Network for Rotation Invariant Face Detection. In: Tan, T., Shi, Y., Gao, W. (eds.) ICMI 2000. LNCS, vol. 1948, pp. 237–244. Springer, Heidelberg (2000)
3. Tao, M., Pen, Z., Xu, G.: The Feature of Human Skin Color. *Journal of Software* 12(7) (2001)
4. Eckhorn, R., Reitboeck, H.J., Arndt, M., Dicke, P.: Feature Linking via Synchronization among Distributed Assemblies: Simulations of Result from Cat Visual Cortex. *Neural Computation* 2, 293–307 (1990)
5. Johnson, J.L.: Pulse-Coupled Neural Nets: Translation, Rotation, Scale, Distortion, and Intensity Signal Invariance for Images. *Applied Optics* 33(26), 6239–6253 (1994)
6. Johnson, J.L., Padgett, M.L.: PCNN Models and Applications. *IEEE Trans. Neural Networks* 10(3), 480–498 (1999)
7. Kuntimad, G., Ranganath, H.S.: Perfect Image Segmentation Using Pulse Coupled Neural Networks. *IEEE Trans. Neural Networks* 10, 591–598 (1999)

An Adaptive Hybrid Filtering for Removing Impulse Noise in Color Images

Xuan Guo¹, Baoping Guo², Tao Hu¹, and Ou Yang¹

¹ College of Optoelectronics Science and Engineering
Huazhong University of Science and Technology
Wuhan 430074, China

² Institute of Optoelectronics, Shenzhen University
Shenzhen 518060, China

Abstract. An adaptive hybrid filter combining a group of sigma vector median filters with different thresholds with a filter based on neuro-fuzzy system is proposed for color image processing. The first subunit of the proposed filter is six sigma vector median filters, their outputs are used as optimum initial points to input the second subunit constituted by a simple Sugeno-type neuro-fuzzy system, and then the optimized result is obtained from the output of the second subunit. The parameters of the neuro-fuzzy model are automatically tuned and fixed by a learning method based on genetic algorithm. The results have indicated that the design of the proposed hybrid filter has met the requirement of removing impulse noise and preserving details. The proposed filter performs better than other filters.

Keywords: Image filtering, Fuzzy neural network, Noise attenuation, Genetic algorithm, Adaptive technique.

1 Introduction

Impulse noise can severely affect subsequent image processing such as edge detecting, image segmentation, object perception and etc, therefore impulse noise filtering, which should have noise-smoothing and detail-preserving qualities, is an essential part of many image processing systems [1-3]. Nonlinear methods are often adopted to restore color images distorted by the impulse noise because of their ability to attenuate impulse noise without degrading the image structure. Generally, insufficient filtering or excessive filtering may result in a nonoptimal filtering output. The vector median filter (VMF) [4], the basic vector directional filter (BVDF) [13] and the directional distance filter (DDF) [14] are some well-known nonlinear filters, but their main drawback is blurring edges and fine details. For that reason, many filters, such as weighted median filters [5-7], filters based on the switching concept [8-9], have been continuously proposed to improve the nonlinear filter. At the same time, another method – ANFIS (Adaptive Neural-Fuzzy Inference System) - has become a complete and powerful framework for image processing. Adaptive neuro-fuzzy reasoning yields one more choices to design very effective nonlinear filters [10-11]. ANFIS able to automatically generate the optimized rulebase is a very useful method to obtain better results.

Lukac et al. proposed a new adaptive filter, Adaptive Sigma Vector Median Filter (ASVMF) [12], constituted by an efficient switching rule between filter output and no filtering, and order-statistic concepts are used in the strategy. Although ASVMF exhibits better performance than many other filters mentioned in Lukac’s paper, there is some space to be improved. Hence a hybrid filter combining a group of sigma vector median filters (SVMF) [12] with a Sugeno-type Neuro-fuzzy system is proposed in this paper. By means of neural network’s adaptive ability, the optimized result is obtained while the complicated nonlinear mapping relation is met.

The rest of the paper is organized as follows. Section 2 introduces the proposed filter model. Experimental results are exhibited in section 3. Finally, section 4 presents some conclusions.

2 The Proposed Filter Structure

The main idea of proposed filter is that the Neuro-fuzzy system uses outputs from a group of SVMFs as optimum initial points, and then better results are achieved by training the neural network. It is shown in Fig.1 that the proposed hybrid filter is composed of two cascaded subunits. The first subunit is six SVMFs, and each of them has a different switching threshold, which aims at inputting different filtering results between VMF operation and identity operation. $SVMF^{(n)}$, $n=0, \dots, 5$, represents the n -th SVMF, which has a threshold of $Tol^{(n)}$. The second subunit is a simple Neuro-fuzzy system in order to further obtain an optimized output.

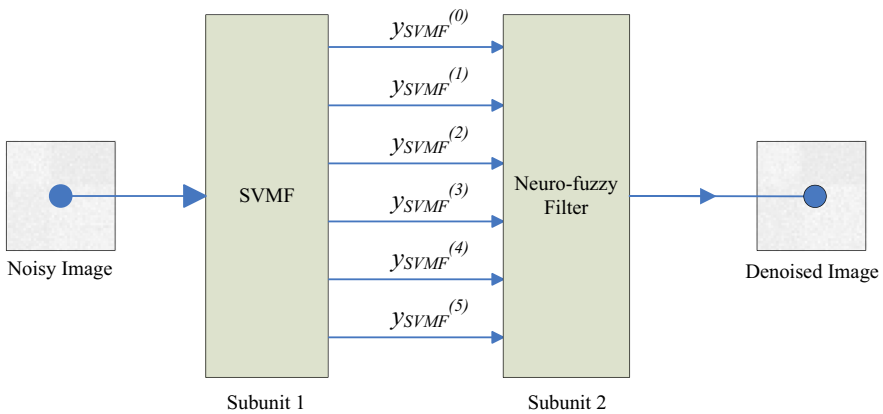


Fig. 1. Structure of the proposed hybrid filter

2.1 Sigma Vector Median Filter

y_{SVMF} which the SVMF outputs is defined as follows:

$$y_{SVMF} = \begin{cases} x_{(1)} & L_c \geq Tol \\ x_c & otherwise \end{cases} \tag{1}$$

$$L_{(1)} = \sum_{i=1}^N \|x_{(1)} - x_i\| \tag{2}$$

where N is the number of the samples in a filtering window, $x_{(1)}$ is the vector median, $L_{(1)}$ is the aggregated distance calculated by (2), and L_c denotes the distance measure associated with the central pixel x_c . The switching concept depends on the threshold Tol . If L_c is larger or equal to the threshold Tol , then x_c is replaced with $x_{(1)}$; otherwise x_c is kept unchanged. The threshold Tol is given by

$$Tol = L_{(1)} + \frac{\lambda * L_{(1)}}{N - 1} \tag{3}$$

Where λ is the tuning parameter used to adjust the smoothing properties of the SVMF, and the value of λ ranges from 0 to $(N-1)(N-2)$. A SVMF with a smaller λ is close to VMF operation; and a SVMF with a larger λ performs no filtering operation. As a consequence, the SVMF can be appropriately varied to trade off between noise suppression and detail preservation. For a filtering window with the size of 3×3 pixels, $N=9$ and $\lambda \in [0,56]$. $y_{SVMF}^{(n)}$, $n=0,1,\dots,5$, is defined as

$$y_{SVMF}^{(n)} = \begin{cases} x_{(1)} & L_c \geq Tol^{(n)} \\ x_c & otherwise \end{cases} \tag{4}$$

$$Tol^{(n)} = L_{(1)} + \frac{\lambda^{(n)} * L_{(1)}}{8} \tag{5}$$

where $Tol^{(n)}$ is the threshold of SVMF⁽ⁿ⁾, $y_{SVMF}^{(n)}$ is the output of SVMF⁽ⁿ⁾.

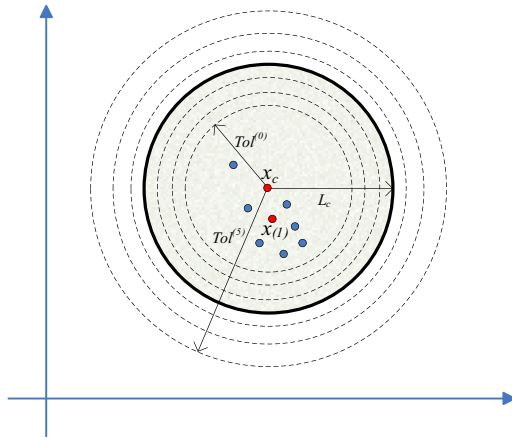


Fig. 2. The scheme of the sigma vector median filtering in the two-dimensional case

Fig.2 indicates the radius of influence with each $Tol^{(n)}$, which decide whether the central pixel x_c is replaced by $x_{(1)}$. In fact, SVMF⁽⁰⁾ is equivalent to a VMF, and SVMF⁽⁵⁾ is much the same as no filtering.

Table 1. Dependence of the $\lambda^{(n)}$ on the parameter n

n	0	1	2	3	4	5
$\lambda^{(n)}$	0	2	5	12	28	50

Tab.1 shows that the $\lambda \in [0, 56]$ is divided into six intervals, and of course other principles of dividing may also exist. The selection depends on it that the fine structure in filtering window should be better discriminated so as to make up the drawback of VMF.

2.2 Sugeno-type Neuro-fuzzy System

The second subunit is a classic sugeno-type neuro-fuzzy system with six inputs and one output. Let X_i denote the inputs of the neuro-fuzzy system and Y denote its output. Each input has two generalized bell-type membership functions (6), and it means that $64 (2^6)$ rules constitute the whole rulebase. The Euclidean distance between each input and the central pixel is used as the input of fuzzy sets (7). The output has a linear membership function of the inputs (8). Respectively, O_k and w_k are defined as (9) (10),

$$M_{ij}(X_i) = \frac{1}{1 + \left| \frac{u_i - a_{ij}}{b_{ij}} \right|^{2c_{ij}}} \tag{6}$$

$$u_i = \|x_c - X_i\| \tag{7}$$

$$Y = \frac{\sum_{k=1}^{64} w_k O_k}{\sum_{k=1}^{64} w_k} \tag{8}$$

$$O_k = d_{k1}X_1 + d_{k2}X_2 + d_{k3}X_3 + d_{k4}X_4 + d_{k5}X_5 + d_{k6}X_6 + d_{k7} \tag{9}$$

$$w_k = \begin{cases} M_{11}(X_1)M_{21}(X_2)M_{31}(X_3)M_{41}(X_4)M_{51}(X_5)M_{61}(X_6) & k=1 \\ M_{11}(X_1)M_{21}(X_2)M_{31}(X_3)M_{41}(X_4)M_{51}(X_5)M_{62}(X_6) & k=2 \\ \vdots \\ M_{12}(X_1)M_{22}(X_2)M_{32}(X_3)M_{42}(X_4)M_{52}(X_5)M_{62}(X_6) & k=64 \end{cases} \tag{10}$$

where $i=1,2,\dots,6, j=1,2,$ and $k=1,2,\dots,64$. The values of a, b, c, d are determined by training.

Here, how the standard sugeno-type neural network works will be no longer repeated. Of course, the ability to remove noise can be improved by taking account more inputs. But the redundant inputs not only increase both the training time and the hardware overhead, but complicate the whole neural network.

2.3 Training Neuro-fuzzy System

The internal parameters (a , b , c , d) of the neuro-fuzzy system are tuned by training, so that its outputs approach to the noise-free training image. The leaning method based on the genetic algorithm is adopted to achieve optimized results. The noise-free target image is a 128×128 pixel Lena color image, and the noisy training image is obtained by corrupting the target image by 10% probability impulse noise. The impulse noise can be defined as:

$$x_{ik} = \begin{cases} o_{ik}, & \text{with probability } 1-p\% \\ v_{ik}, & \text{with probability } p\% \end{cases} \quad (11)$$

where x_{ik} are the pixels of the noise image, o_{ik} represents the noise-free samples and v_{ik} represents the pixels corrupted by impulsive noise,. In the noise model, the value of v_{ik} can take on all integers from 0 to 255 with equal probability, and the contamination of three components ($k=1, 2, 3$) in the color image is uncorrelated.

A genetic algorithm starts with a randomly generated population of individuals and produces the subsequent populations by mean of reproduction, crossover and mutation operators. The fitness of each individual corresponding neuro-fuzzy system performance is measured. The individuals having the best fitness have more chance to reproduce their descendants. The learning process stops when an expected value of fitness has been obtained.

We have set the parameters of genetic learning as follows: individual population 20, elite count 2, crossover fraction 0.8, mutation function is Gaussian, stall generation 50, and the fitness function is based on the mean-square error (MSE), given by

$$MSE = \frac{\sum_{i=1}^N \sum_{k=1}^m (y_{ik} - o_{ik})^2}{N * m} \quad (12)$$

where y_{ik} is the outputs of the neuro-fuzzy system, o_{ik} are the pixels of the noise-free image and $m=3$ denotes 3 channels of color image, $N=128 \times 128$ is the number of the training pixels.

3 Experimental Results

Two test images, Lena and Peppers (both of them 256×256), were employed to evaluate the performance of the proposed filter. A quantitative evaluation of the filter can be given by estimating the mean-square error (MSE) of the images processed with the proposed and some other filter. The output images of some operators for the Lena and Peppers color image corrupted by impulse noise with 10% probability are shown in Fig 3.



Fig. 3. a) Noisy image with impulses (10%), b) The output of the VMF, c) The output of the BVDF, d) The output of the DDF, e) The output of the ASVMF, f) The output of the proposed filter

It is observed from Tab 2 and Tab 3 that the proposed technique largely outperforms other filters. The proposed method has also been compared with ASVMF. It is seen that the fuzzy method yields better results, especially for relatively low noise probability. Interestingly, the ASVMF performs worse than the VMF in highly corrupted images.

Table 2. Filtering results achieved using test image Lena

Lena	5%	10%	15%	20%
Noisy	444.36	891.29	1331.34	1782.44
VMF	51.59	65.15	87.59	122.11
BVDF	63.35	87.52	128.08	183.87
DDF	56.59	74.24	104.33	145.42
ASVMF	38.16	83.81	158.73	260.27
proposed	15.45	28.91	51.58	73.62

Table 3. Filtering results achieved using test image Peppers

Peppers	5%	10%	15%	20%
Noisy	513.47	1005.94	1490.73	2007.81
VMF	37.76	55.50	80.72	130.88
BVDF	77.70	129.84	182.05	288.87
DDF	46.18	70.66	102.09	168.17
ASVMF	32.36	86.68	164.95	295.09
proposed	17.87	32.55	58.67	82.10

4 Conclusions

An adaptive hybrid filter for removing impulse noise in color images has been presented. By a GA learning method, the proposed filter is able to learn from training examples the reasonable rulebase, and the network maps the inputs variables to the optimized output variable. It is very effective to remove noise without degrading the image details. The experimental results show that the proposed filter provides better performance than other filters.

References

1. Plataniotis, K.N., Venetsanopoulos, A.N.: *Color Image Processing and Applications*. Springer, Heidelberg (2000)
2. Lukac, R., Smolka, B., Martin, K., Plataniotis, K.N., Venetsanopoulos, A.N.: Vector Filtering for Color Imaging. *IEEE Signal Process. Mag. Spec. Issue on Color Image Processing* 22, 74–86 (2005)
3. Zheng, J., Valavanis, K.P., Gauch, J.M.: Noise removal from color images. *J. Intelligent and Robotic Syst.* 7, 257–285 (1993)
4. Astola, J., Haavisto, P., Neuvo, Y.: Vector Median Filters. *Proceeding of the IEEE* 78, 678–689 (1990)

5. Lukac, R., Smolka, B., Plataniotis, K.N., Venetsanopoulos, A.N.: Selection Weighted Vector Directional Filters. *Comput. Vision Image Understand. Special Issue on Color for Image Indexing and Retrieval* 94, 140–167 (2004)
6. Lucat, L., Siohan, P., Barba, D.: Adaptive and Global Optimization Methods for Weighted Vector Median Filters. *Signal Process.: Image Commun.* 17, 509–524 (2002)
7. Viero, T., Oistamo, K., Neuvo, Y.: Three-dimensional Median Related Filters for Color Image Sequence filtering. *IEEE Trans. Circuits Syst. Video Technol.* 4, 129–142 (1994)
8. Lee, J.S.: Digital Image Smoothing and The Sigma Filter. *Computer Vision Graph. Image Process* 24, 255–269 (1983)
9. Eng, H.L., Ma, K.K.: Noise Adaptive Soft-switching median filters. *IEEE Trans. Image Process.* 10, 242–251 (2001)
10. Russo, F.: Hybrid Neuro-fuzzy Filter for Impulse Noise Removal. *Pattern Recognition* 32, 1843–1855 (1999)
11. Yuksel, M.E.: A Median/ANFIS Filter for Efficient Restoration of Digital Images Corrupted by Impulse Noise. *Int. J. Electron. Commun.* 60, 628–637 (2006)
12. Lukac, R., Smolka, B., Plataniotis, K.N., Venetsanopoulos, A.N.: Vector Sigma Filters for Noise Detection and Removal in Color Images. *J. Vis. Commun. Image R.* 17, 1–26 (2006)
13. Trahanias, P.E., Venetsanopoulos, A.N.: Vector Directional Filters: A New Class of Multichannel Image Processing Filters. *IEEE Trans. Image Process.* 2, 524–528 (1993)
14. Karakos, D.G., Trahanias, P.E.: Combining Vector Median and Vector Directional Filters: The Directional Distance Filters. In: *IEEE ICIP Conf.*, pp. 171–174. IEEE Press, New York (1995)

A Multi-Stage Neural Network Model for Human Color Vision

Charles Q. Wu

Stanford Continuing Studies, Stanford University, Stanford, CA 94305, U.S.A.
charlesqwu@126.com

Abstract. The contemporary “Standard Model” for human color vision is a two-stage model: The first stage consists of three types of receptors at the retina of the eye, and the second stage consists of three opponent-color neural channels: Red-Green, Blue-Yellow, and Black-White. In this paper I call upon the phenomena of complementary afterimages and of the “flight of colors” to show that this model is not an adequate explanation for these color phenomena in specific and for color vision in general. To remedy the theoretical inadequacy of the “Standard Model”, I propose a neural stage for color complementarity directly corresponding to our color sensation. Mapping onto the anatomical organization of human visual system, I further suggest that layer 4C in the primary visual cortex is the neural substrate directly responsible for color complementarity in particular and for color appearance (that is, color consciousness) in general.

Keywords: Complementary colors, Opponent colors, Afterimages, Flight of colors, Layer 4C, Primary visual cortex, Color consciousness.

1 Introduction

Color vision is a fascinating subject matter of research in itself. Not only that, studying the color visual system can also be very illuminating – this is because its computational task is a generic computational problem: using three overlapping filters to construct perceptual categories as well as a continuum of sensations. What are the underlying neural circuitry and mechanisms for this marvelous accomplishment? Studying such a system would certainly shed light on understanding human perception and cognition in general.

As Mollon relates, the period of AD 1850-1931 is a “golden age” of color research as quantitative data about human color perception were obtained and theories of color vision were proposed. Two prominent theories for color vision coming out of this period are Young-Helmholtz's trichromatic theory and Hering's opponent-process theory. The basic tenets of the trichromatic theory are that we humans are endowed with three types of color receptors and that all of our color sensations can be conceived as additive outcomes of the responses from these receptors. On the other hand, the opponent-process theory maintains that we have four unique hues (Red, Green, Yellow, and Blue) in our subjective experience of colors and that these unique hues, along with the achromatic colors Black and White, compose three opponent-color

channels: Red-Green, Blue-Yellow, and Black-White. Originally, these two theories of color vision were thought completely incompatible with each other. Nevertheless, subsequent to Helmholtz and Hering, many researchers had attempted to reconcile these two schools of thoughts – among them are von Kries, Mueller, and Ladd-Franklin (see Mollon, 2003). A very influential model coming out such attempts to reconcile the two color theories is the work performed by Hurvich and Jameson (1957) about a half-century ago – as a matter of fact, their model is so influential that it has now become the “Standard Model” in color vision research. As illustrated in Figure 1, this model consists of two stages: The first stage is of three types of photoreceptors responding to short-, medium-, and long-wavelengths (SW, MW, and LW) of the light; and the second stage is of three opponent-color neural channels for Red-Green, Yellow-Blue, and Black-White. As shown in Figure 1, Hurvich and Jameson also postulated the neural connectivity from the first stage to the second one in their model.

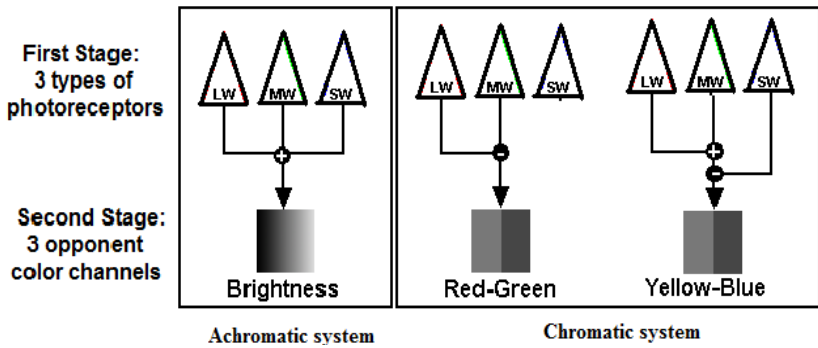


Fig. 1. The contemporary “Standard Model” of human color vision

In this paper, I will demonstrate that the above “Standard Model” is inadequate in explaining certain basic color phenomena – particularly the phenomena of complementary afterimages and of the flight of colors. Following that, I will propose a multi-stage, neuroanatomically-based neural network model for human color vision – in this model, layer 4C of the primary visual cortex (V1) is identified as the neural stage or substrate directly corresponding to color complementarity, color mixing, and color appearance (that is, color consciousness). The theoretical model presented here should be able to serve as a basis for large-scale, neurobiologically-based neural network models that would attempt to relate human color subjective experience and relevant psychophysical data directly with underlying neural mechanisms in our brain.

2 Negative / Complementary Afterimages

If one looks at a Red patch for 30 seconds and then looks at a white background, one will see a colored patch of Cyan – the complementary color of Red. Figure 2 represents three pairs of complementary colors as opposite colors on a color circle:

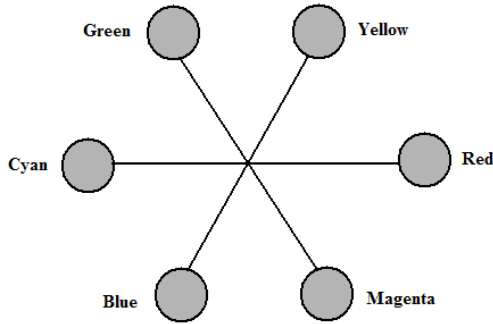


Fig. 2. Complementary colors as opposite-site colors around a color circle

Looking at any color patch around the color circle will produce an afterimage of its complementary color. This phenomenon is known as “negative or complementary afterimage”. (The term “negative afterimage” is usually used only for achromatic images while the term “complementary afterimage” is used only for chromatic images. Nonetheless, as Ladd-Franklin (1929) pointed out, our ordinary word “color” can be used in two senses: In its narrow sense, it means only chromatic colors; whereas in its broad sense, it refers to both chromatic and achromatic colors. Following Ladd-Franklin (1929), throughout the present paper, we will use the word “color” in its broad sense and therefore we will use the terms “complementary afterimage” and “negative afterimage” interchangeably.)

Afterimage is a seemingly simple yet very intriguing visual phenomenon and is always mentioned in textbooks on perception and general psychology (as well as in many popular science books and on many Internet sites). Unfortunately, at the time being, many such books have been propagating an incorrect description about negative afterimages: “After looking at a RED color, you will experience a GREEN afterimage” (e.g., see Wolfe, Kluender, & Dennis, 2005). This description is incorrect because the complementary afterimage of RED is not GREEN but CYAN. Likewise, the complementary afterimage of GREEN is not RED but MAGENTA. This incorrect description about negative afterimages has propagated to its current status partly because of a linguistic twist: The color Cyan is a binary color which many people would use the phrase “Greenish-Blue” (or “Royal Blue”, “Sky Blue”, “Rain-drop Blue”) to describe. In this sense, the above description may just be said as “inaccurate”. However, more fundamentally, this description is incorrect because of its inherent misconception due to using Hering’s opponent-process theory to explain complementary afterimages. We will discuss this issue in more detail below. At any rate, regardless of any theoretical orientation, we should be clear about the fact that the colors exhibited in negative afterimages are complementary colors: Red vs. Cyan, Green vs. Magenta, and Blue vs. Yellow.

Besides this fact, we should also be aware of the following well-known facts about negative afterimages: They are retinotopic, monocular (that is, they usually do not show inter-ocular transfer from an adapted eye to the other eye), and based on visual surface representations. The last fact had been convincingly demonstrated by Shimojo et al. (2004). Their procedure for experimentally showing this aspect of afterimages is

somewhat complicated, but it can readily demonstrated with a simple visual stimulus as shown in Figure 3: After looking at the stimulus in Figure 3(a), one will experience a monocular rivalry sequence of three afterimages shown in Figure 3(b). This is as if the visual system constructs three surface representations out of the original visual stimulus and then the three afterimages resulting from these surface representations compete with each other and generate the perceived monocular rivalry among them. There has been ample visual psychophysical evidence indicating that the visual system decomposes the 2-D visual stimulation into surfaces or layers in order to eventually construct a 3-D representation (see Nakajama and Shimojo, 1990). Here in afterimages, it appears that each surface representation (or layer) is capable of eliciting an afterimage and that the afterimages on different layers could engage in competition with each other and show off in the form of monocular rivalry.

As noted by De Valois and De Valois (1997), two major theories have been offered to explain complementary afterimages: one stemming from the trichromatic theory and the other from the opponent-process theory. According to Mollon (2003), George Palmer (1786) was a pioneer in deriving the idea of human trichromacy and was first to propose a “fatigue” theory for negative afterimages: After staring at a certain color, the corresponding type of receptors in the retina becomes over-exerted and this creates an imbalance among the three types of receptors; on subsequent exposure to light stimulation, the over-exerted (that is, fatigued) receptor type would become less-excited and therefore the color complementary to the original color would be seen. (Formally, the fatigue theory can be expressed in this manner: Assume that the three receptors are R, G, and B; and assume that the R receptor becomes fatigued during the initial visual stimulation and is only partially-excited on subsequent white stimulation; then the perceived afterimage is: $\alpha R + 1.0G + 1.0B = \text{White} - \beta R = \beta R'$, $\beta = 1 - \alpha$, where R, G, and B are the responses in the three receptors; α is the rate of partial-excitation of the R receptor during the afterimage; and R' is the complementary color of R.)

Almost two hundred years later after Palmer's original proposal, the “fatigue” part of his explanation for negative afterimages was given a particular physiological mechanism: photopigment bleaching in the retinal photoreceptors (Brindley, 1957). Even though there seems ample psychophysical and physiological evidence suggesting the existence of photopigment bleaching, the purported causal relation between bleaching and negative afterimages remains elusive. Note that the fatigue explanation depends on the ratios among the three receptor channels – according to it, negative afterimages would appear only with subsequent retinal stimulation. However, negative afterimages also occur on dark backgrounds or when the eyes are closed; and as pointed out by De Valois and De Valois (1997), this is one of the prominent difficulties with the fatigue theory of negative afterimages. Besides, there are also many other aspects of negative afterimages that the fatigue theory does not seem to be able to explain (e.g., see Loomis, 1972 and Shimojo et al., 2001). As we will see below, the fatigue / bleaching theory can not explain color reversals occurred in the “flight of colors” phenomenon and does not seem to hold true even when the stimulus is an intense light.

Hering's opponent-color explanation for negative afterimages is that the adaptation seen in such afterimages is due to mutual interaction within each of the opponent-process channels. However, the problem with this explanation is that the colors

displayed in negative afterimages are not opponent colors at all, but complementary colors – This is what we have already emphasized above. This problem with Hering's explanation was recognized long before – for example, as Ladd-Franklin (1929) (an American woman psychologist who studied with both Helmholtz's and Hering's disciples in Germany) pointed out, Red and Green are yellow-constituting colors, not white-constituting colors. In other words, Red and Green do not cancel each other; and they do not show up in relationship in complementary afterimages.

As pointed out by Pridmore (2008), a half-century ago, authors by and large used complementary colors to describe negative afterimages and there have been many detailed psychophysical studies of afterimages indicating the complementary nature of negative afterimages. The transition from complementary colors to opponent colors in describing negative afterimages happened thanks to Hurvich and Jameson's (1957) influential work. If one examines their experiments and the relevant theoretical interpretations, one would see that their work is indeed built up the *a priori* assumption of four unique hues conceived by Hering. Therefore, though very influential indeed, their work does not change the nature that the opponent-process theory is an inadequate explanation for complementary afterimages. Of course, the opponent-process theory may still be true and valuable for interpreting certain other color-related vision aspects, but for the complementary signature of negative afterimages the supposed opponent stage is certainly a sham. Below I will call upon another color phenomenon to prove beyond doubt that the opponent-process stage as speculated by Hering and further developed by Hurvich and Jameson can not possibly be the neural substrate for afterimages in particular and for many other color perceptual phenomena in general.

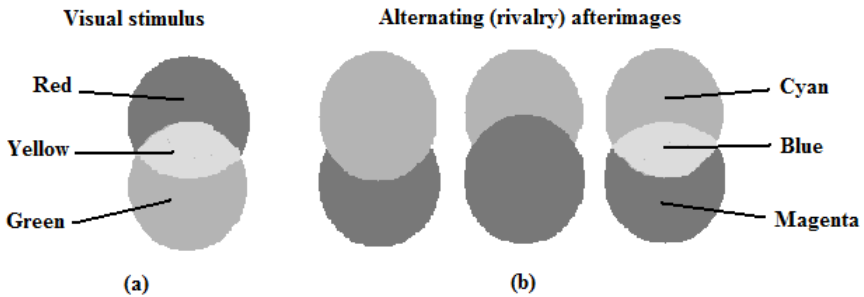


Fig. 3. Visual surface representations in afterimages: The visual stimulus in (a) can elicit the various afterimages (which alternate among themselves in time) illustrated in (b)

3 The Flight of Colors (FOC)

If one looks at an intense white light (e.g., the bright sun in a clear sky) for a short duration and then closes his/her eyes, one will experience a sequence of colors. This phenomenon is known as the “flight of colors (FOC)”. According to Barry and Housfield (1934), this phenomenon has been known since Aristotle and has been described by such inquisitive minds as Newton, Goethe, De Vinci, among many others. What colors

do appear in FOC? Pondering over this question by itself may just be curiosity-satisfying. But once we think about this question with a theoretical interest, particularly in relation to the trichromatic theory versus the opponent-process theory controversy, we can realize that this phenomenon may serve as a critical test of these theories.

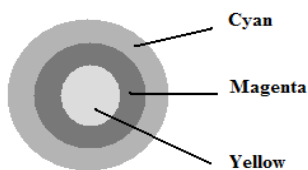


Fig. 4. The spatial organization and the main colors in the flight of colors

Many investigators had described the colors in FOC, but it appears that they had used color names rather inconsistently – for instance, some authors had used the term “Greenish-Blue” or just “Blue” for the color that we would now call “Cyan”. Therefore I have made many observations staring directly and briefly (< 1 second) at Northern California’s bright sun and then closing or opening my eyes to experience FOC. The main results of my observations can be summarized as follows: (1) As illustrated in Figure 4, the FOC afterimage usually consists of one or two rings surrounding an inner disk; (2) With closed eyes, the predominant colors occurring in FOC are Cyan, Magenta, and Yellow; (3) With open eyes and projecting FOC onto a white background, the colors would immediately change to their complementary ones – this is possibly the reason why Helmholtz described the colors in FOC as Red, Green, and Blue; (4) Other colors also occur in FOC but they mainly result from color filling-in and mixing in parts of the afterimage (e.g., Magenta from a surrounding ring would intrude into the central disk of Yellow to produce an orange color).

As listed above, an important observation about FOC is that during FOC, opening eyes to look at a white background would immediately yield “hue reversal” – that is, the colors in FOC would instantly change to their complementary ones. This fact renders the fatigue explanation completely implausible for this phenomenon: “Fatigue” entails that there should be a refractory period to recover from an adapted state to the original, neutral point or to the opposite direction of adaptation, but the current fact about FOC indicates that this is not the case. Hence, we need to abandon the hypothesis of photoreceptor fatigue (or bleaching) altogether in explaining afterimages and the flight of colors.

Now we see that FOC is more consistent with the trichromatic theory, yet the fatigue explanation as originally offered by the trichromacy pioneer George Palmer is also inadequate. Therefore, we are left with the only possibility of tri-modal color cardinals corresponding to three primary colors (Red, Green, and Blue) and their complementary ones (Cyan, Magenta, and Yellow) at a neural level instead of at the photoreceptor level. Below I will go on to suggest that this happens in layer 4C of the primary visual cortex (that is, visual cortical area V1).

4 A Multi-Stage Model for Color Vision with a Cortical Stage for Complementary Colors

Before going on to present a new multi-stage model for color vision, we will need to briefly review the relevant neuroanatomy. Figure 5 depicts the first few stages of the primate (including the humankind) visual system: When visual stimulus falls upon the retina of the eye, the photoreceptors there catch photons and convert them into electrical signals; and then the neural circuitry in the retina converges such signals and, through the LGN (lateral geniculate nucleus) of the thalamus, conveys these signals onto layer 4C in the primary visual cortex (which is also known as the striate cortex, visual cortical area V1, and Brodmann's area 17).

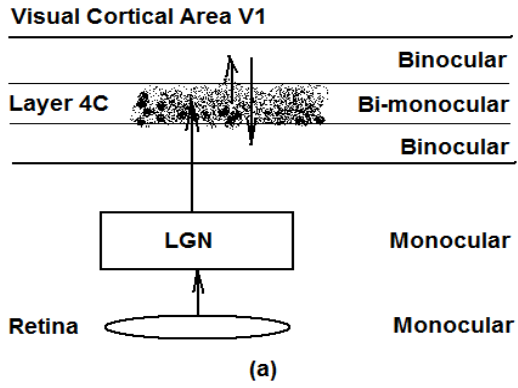


Fig. 5. The anatomical organization of the early stages of the primate (including the human-kind) visual system

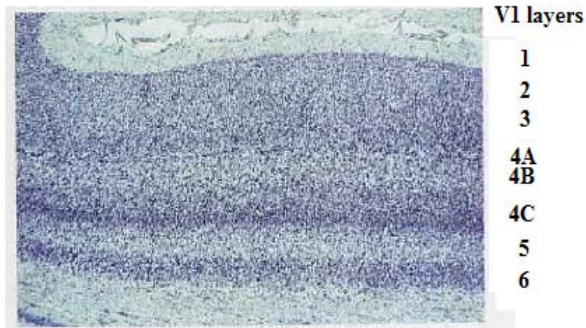


Fig. 6. A Nissl-stained section showing the laminar organization of the primary visual cortex (also known as, visual cortical area V1)

An important anatomical, as well as physiological, feature concerning the early parts of the primate visual system is that layer 4C is the ONLY cortical stage where cells are predominantly monocular. However, unlike in one retina or one layer of one

LGN where all the neurons are completely monocular – receiving their inputs solely from one eye – those monocular neurons receiving inputs from the left eye and others receiving inputs from the right eye co-exist in the same place in layer 4C, in the fashion of the so-called “ocular dominance columns” (see Horton, 2006). Because of this co-existence in layer 4C of both LE (left eye) monocular neurons and RE (right eye) counterparts, I suggest that this layer be more appropriately referred to as a “bi-monocular” layer.

Another prominent anatomical feature of layer 4C is that there is a gradient of cell density through the depth of this layer, as schematically illustrated in Figure 5 (see Lund et al., 1995). The neuroanatomical technique of Nissl stain can be used to show the cell bodies of neurons in a small tissue of the brain. The above-mentioned feature can be readily seen in a Nissl-stained section, cutting from the pia to the white matter of the cortical sheet, of V1 – as shown in Figure 6, careful examination of layer 4C there reveals that there is a gradual increase of cell density from the top of layer 4C to its bottom. Furthermore, this anatomical feature may be correlated with a continuum of neurons' contrast sensitivities through the depth of layer 4C (see Lund et al., 1995).

We now know that the site for negative afterimages is neural, retinotopic, monocular, and capable of representing visual surfaces (layers). We also know that layer 4C in V1 is the only cortical substrate meeting these features / requirements. Therefore, I suggest that layer 4C is the neural substrate for negative afterimages and the flight of colors – and more importantly, for color complementarity. Negative afterimages and the flight of colors are just phenomenal revelations of color complementarity. The complementarity feature is a real signature of some underlying neural mechanism directly responsible for color consciousness. Therefore, we now have a three-stage model for human color vision with layer 4C of V1 incorporated in the model as a neural stage for color complementarity and for color consciousness.

Color complementarity implies that there must be some neural mechanism performing “subtraction” in the sense of subtractive color mixing. This is precisely because of the nature of complementary colors as expressed in the relation: $C' = \text{White} - C$, where C and C' are a pair of complementary colors. The fact that subtractive color mixing occurs in layer 4C also suggests that additive color mixing, the other side of the same coin, must also be there in the same neural substrate. Then, what could be the possible neural mechanism for additive color mixing?

In computational neuroscience, the problem of combining features along multiple dimensions to create a coherent representation over a distributed neural network is known as the “binding problem”. It has long been suggested that the solution to the binding problem is neural synchronization (see von der Malsburg, 1994), and there has been a growing body of physiological evidence supporting this conclusion (e.g., see Engel and Singer, 2001). In color vision, we have exactly the same type of binding problem: How to bind the outputs from the three color cardinals to produce any color in the whole gamut of perceivable colors? In this regard, I suggest that neural synchronization is indeed the physiological mechanism for combining outputs from cardinal color cells to create a whole spectrum of color sensations. In this conception, the numbers of cells firing together, instead of their firing rates averaged over certain

time windows, are physiological measures of color components and constitute the weights in the color mixing equation: $C = \alpha R + \beta G + \gamma B$, $\alpha + \beta + \gamma = 1$, where C is any color; R , G , and B are the three cardinal colors; α , β , and γ are the weights for color C along the color cardinals.

An important piece of evidence supporting human trichromacy is the color-matching functions derived by Helmholtz and his colleagues (see Mollon, 2003). Such color-matching functions were originally thought to reflect the sensitivities of the three types of photoreceptors in the retina. Currently we already know that such functions do not directly correspond to the sensitivities of the three types of cones in the retina. Here, as we can see, the color-matching functions are actually the measures of cortical neurons in layer 4C – rather than the photoreceptors in the retina. Similarly, the hue cancellation technique employed by Jameson and Hurvich (1957) also utilizes the nature of color mixing. Therefore, even though these investigators claimed their results as evidence supporting Hering's opponent-color theory, in reality they were conducting psychophysical measures corresponding to the neural stage of color complementarity in layer 4C of V1.

A special case of color mixing is color transparency where two colors are not really mixed but one is seen in front of the other – that is, the two colors are seen in two separate depth planes. A variant of Metelli's relation for perceptual transparency is also a linear formula: $C = \alpha C1 + \beta C2$, $\alpha + \beta = 1$, where $C1$ and $C2$ are two transparent colors and C is the color of their overlapping area. From our perspective, we can assert that color transparency occurs as the result of two sub-layers of neurons in layer 4C firing synchronously. Just like the situation of normal additive color mixing, the weights in the above Metelli's formula are the numbers of cells firing together at these sub-layers. Of course, color transparency depends on the spatial organizations (or patterns) of the two overlapping colors, and such spatial factors affect the segregation of neurons into sub-layers in layer 4C – corresponding to our perceptual “color scission” (Metelli, 1974). Of course, the direct correspondence between color scission and the underlying neural circuitry remains to be discovered through neuroanatomical and neurophysiological research.

5 Conclusions

(1) The opponent-process color theory as conceived by Hering and further developed by Jameson and Hurvich is not an adequate explanation for complementary afterimages and the flight of colors; (2) Hue reversal in the flight of colors implies that color complementarity occurs at a neural level instead of as some retinal “fatigue” (photopigment bleaching) process; (3) Mapping onto the anatomical organization of the primate visual system, it is evident that layer 4C in V1 is the neural substrate directly responsible for negative afterimages and for color complementarity; (4) The proposed neural substrate also holds the mechanism for color mixing, transparency, filling-in, and fading-out: These perceptual phenomena correspond to synchronization or dys-synchronization of clusters of neurons along the color cardinals. In short, layer 4C of the primary visual cortex is the neural substrate for color consciousness.

References

1. Barry Jr., H., Housfield, W.A.: Implications of the Flight of Colors. *Psychological Review* 41, 300–305 (1934)
2. Brindley, G.S.: The Discrimination of After-images. *Journal of Physiology* 147, 194–203 (1957)
3. De Valois, R.L., De Valois, K.K.: A Multi-stage Color Model. *Vision Research* 33, 1053–1065 (1993)
4. De Valois, R.L., De Valois, K.K.: Neural Coding of Colour. In: *Readings on Color. Science of Color*, vol. 2, pp. 94–140. Bradford Books/MIT Press, Cambridge (1997)
5. Engel, A.K., Singer, W.: Temporal Binding and the Neural Correlates of Sensory Awareness. *Trends in Cognitive Sciences* 5, 16–25 (2001)
6. Ladd-Franklin, C.: *Colour and Colour Theories*. Harcourt Brace, New York (1929)
7. Horton, J.C.: Ocular Integration in the Human Visual Cortex. *Canadian J. Ophthalmology* 41, 584–593 (2006)
8. Hurvich, L.M., Jameson, D.: An Opponent-process Theory of Color Vision. *Psychological Review* 64, 384–404 (1957)
9. Loomis, J.M.: The Photopigment Bleaching Hypothesis of Complementary Afterimages: A psychophysical test. *Vision Research* 12, 1587–1594 (1972)
10. Lund, J.S., Wu, Q., Levitt, J.B.: Visual Cortex Cell Types and Connections: Anatomical Foundations for Computational Models of Primary Visual Cortex. In: *The Handbook of Brain Theory and Neural Networks*, 1st edn., pp. 1016–1021. Bradford Books/MIT Press, Cambridge (1995)
11. Metelli, F.: The Perception of Transparency. *Scientific American* 230, 91–98 (1974)
12. Mollon, J.D.: The Origins of Modern Color Science. In: *Color Science*, pp. 1–39. Optical Society of America, Washington DC (2003)
13. Nakayama, K., Shimojo, S.: Towards a Neural Understanding of Visual Surface Representation. In: *The Brain*, vol. 55, pp. 911–924. Cold Spring Harbor Laboratory, New York (1990)
14. Palmer, G.: *Théorie de la Lumière, Applicable aux arts, et Principalement à la Peinture*. Hardouin et Gattey, Paris (1786)
15. Shimojo, S., Kamitani, Y., Nishida, S.: Afterimage of Perceptually Filled-in Surface. *Science* 31, 1677–1680 (2001)
16. Pridmore, R.W.: Chromatic induction: Opponent Color or Complementary Color Process? *Color Research and Application* 33, 77–81 (2008)
17. von der Malsburg, C.: *The Correlation Theory of Brain Function*. Departmental Technical Report, Cogprints (1981)
18. Wolfe, J.M., Kluender, K.R., Levi, D.M.: *Sensation and Perception*. Sinauer Associates, Sutherland (2005)

Lead Field Space Projection for Spatiotemporal Imaging of Independent Brain Activities

Huiling Chan¹, Yong-Sheng Chen¹, Li-Fen Chen^{2,3,*}, Tzu-Hua Chen¹,
and I-Tzu Chen¹

¹ Department of Computer Science, National Chiao Tung University, Hsinchu,
Taiwan

² Institute of Brain Science, National Yang-Ming University, Taipei, Taiwan

³ Department of Medical Research and Education, Taipei Veterans General Hospital,
Taipei, Taiwan

Abstract. Magnetoencephalography and electroencephalography are non-invasive instruments that can record magnetic fields and scalp potentials, respectively, induced from neuronal activities. The recordings are superimposed signals contributed from the whole brain. Independent component analysis (ICA) can provide a way of decomposition by maximizing the mutual independence of separated components. Beyond the temporal profile and topography provided by ICA, this work aims to estimate and map the cortical source distribution for each component. The proposed method first constructs a source space using lead field vectors for vertices on the cortical surface. By projecting the specified components to this source space, our method provides the corresponding spatiotemporal maps for these independent brain activities. Experiments using simulated brain activities clearly demonstrate the effectiveness and accuracy of the proposed method.

1 Introduction

Independent component analysis (ICA) is a blind source separation technique that can decompose multi-channel signals into mutually independent components (ICs) [1]. Recently, it has been widely used for analyzing magnetoencephalographic (MEG) and electroencephalographic (EEG) signals [2,3], particularly for removing artifacts based on its independence assumption [4,5,6]. Once the noisy components are removed, others can be mixed to reconstruct the measurements with higher signal-to-noise ratio for further analysis. ICA has two limitations. First, the number of ICs is less than or equal to the number of sensors. Second, conventional ICA can only provide the topography for each component, which is the weighting distribution at sensor level. Therefore, ICA per se is insufficient for mapping cortical source distributions of brain activities.

One commonly-used way to obtain the cortical source distributions of separated ICs is to apply source imaging techniques to the reconstructed measurements without the interference of artifact components [7,8]. There exist in

* Corresponding author.

the literature a wide selection of source imaging techniques, including weighted minimum norm (WMN), dipole fitting, MUSIC, and beamformer-based methods. Recently Tsai et al. [9] proposed another ICA method, called the electromagnetic spatiotemporal independent component analysis (EMSICA), that can simultaneously obtain spatiotemporal ICs and their corresponding cortical source distributions. This method utilizes the Bayesian statistical framework for imaging independent brain activities under physiological source constraints. Unfortunately, the number of unknown parameters in EMSICA is much more than that in conventional ICA and may thus cause higher difficulties and instability when solving the unmixing matrix.

In this work we propose a new source imaging method that directly projects independent components to cortical source space obtained from lead field vectors. It provides an intuitive and efficient solution for analyzing specified independent components. In the rest of this paper, we will describe the proposed algorithm in detail, demonstrate its feasibility by three experiments using simulated brain activities, and draw the conclusions of this work.

2 Methods

2.1 Forward Model

We construct the source space of independent components from the lead field vectors located at all of the vertices on the cortical surface. The lead field vector $\mathbf{l}_\theta \in \mathbb{R}^N$ indicates how a unit dipole with parameters $\theta = \{\mathbf{r}, \mathbf{q}\}$ contributes to the MEG/EEG sensor array:

$$\mathbf{l}_\theta = \mathbf{G}_\mathbf{r} \mathbf{q} , \quad (1)$$

where $\mathbf{G} \in \mathbb{R}^{N \times 3}$ is the gain matrix describing the sensibility of N MEG/EEG sensors to the current dipole located at $\mathbf{r} \in \mathbb{R}^3$ with orientation $\mathbf{q} \in \mathbb{R}^3$ [10][11]. The MEG/EEG measurements $\mathbf{m}(t) \in \mathbb{R}^N$ recorded at time t is composed of D time-varying dipoles:

$$\mathbf{m}(t) = \mathbf{L} \mathbf{s}(t) + \mathbf{n}(t) , \quad (2)$$

where $\mathbf{L} = [\mathbf{l}_{\theta_1} \mathbf{l}_{\theta_2} \dots \mathbf{l}_{\theta_D}]$ is the lead field matrix, $\mathbf{s}(t) = [s_1(t) s_2(t) \dots s_D(t)]^T$ is the time-varying source activities, and $\mathbf{n}(t)$ is the additive noise.

2.2 Source Imaging of Independent Components

ICA can separate the measurements $\mathbf{m}(t)$ into K statistically independent components $\mathbf{x}(t)$:

$$\mathbf{m}(t) = \mathbf{A} \mathbf{x}(t) , \quad (3)$$

where $\mathbf{A} = [\mathbf{a}_1 \mathbf{a}_2 \dots \mathbf{a}_K] \in \mathbb{R}^{N \times K}$ is a mixing matrix that compounds the K independent components $\mathbf{x}(t) = [x_1(t) x_2(t) \dots x_K(t)]^T \in \mathbb{R}^K$ into the measurements $\mathbf{m}(t)$ [1]. Each column vector \mathbf{a}_i in the mixing matrix \mathbf{A} represents the activity distribution on the device sensors corresponding to the i -th component $x_i(t)$, $i = 1, \dots, K$. From another aspect, Eq. (3) can be written as

$$\mathbf{x}(t) = \mathbf{W}^T \mathbf{m}(t) , \quad (4)$$

where $\mathbf{W} = [\mathbf{w}_1 \mathbf{w}_2 \dots \mathbf{w}_K]$ is an unmixing matrix. Each column vector \mathbf{w}_i is a spatial filter for extracting the corresponding component $x_i(t)$ from the measurements $\mathbf{m}(t)$.

Recently, Tsai et al. assume that the K spatiotemporal independent components originate from brain activities at P locations and the matrix $\mathbf{B} \in \mathbb{R}^{P \times K}$ describes this linear relationship [9]:

$$\mathbf{s}(t) = \mathbf{B}\mathbf{x}(t) \quad , \quad (5)$$

where $\mathbf{B} = [\mathbf{b}_1 \mathbf{b}_2 \dots \mathbf{b}_K]$, each column vector \mathbf{b}_i in \mathbf{B} represents the cortical source distribution of the i -th component, $x_i(t)$, $i = 1, \dots, K$. By substituting Eq. (5) into the forward model, Eq. (2) becomes

$$\mathbf{m}(t) = \mathbf{L}\mathbf{s}(t) = \mathbf{L}\mathbf{B}\mathbf{x}(t) \quad . \quad (6)$$

Moreover, by substituting $\mathbf{m}(t)$ with Eq. (6), Eq. (4) becomes

$$\mathbf{x}(t) = \mathbf{W}^T \mathbf{m}(t) = \mathbf{W}^T \mathbf{L}\mathbf{B}\mathbf{x}(t) \quad . \quad (7)$$

Without loss of generality, we assume that the set of $\mathbf{x}(t)$ during a long enough period of time spans the whole space of \mathbb{R}^K . Therefore, the $K \times K$ matrix $\mathbf{W}^T \mathbf{L}\mathbf{B}$ is the identity matrix:

$$\mathbf{W}^T \mathbf{L}\mathbf{B} = \mathbf{I} \quad . \quad (8)$$

The cortical source distribution \mathbf{B} can be derived from

$$\mathbf{B} = (\mathbf{W}^T \mathbf{L})^+ \quad , \quad (9)$$

where the $+$ mark denotes the pseudo-inverse operator and can be performed through singular value decomposition. The \mathbf{W} and \mathbf{L} in Eq. (9) can be obtained from the ICA and the forward model, respectively.

In summary, ICA separates the independent component $x_i(t)$ in sensor space. The proposed method maps independent component $x_i(t)$ into the corresponding source distribution $\mathbf{b}_i(t)$ on the cortical surface through the unmixing matrix \mathbf{W} and the lead field matrix \mathbf{L} .

3 Experiments

To evaluate the performance of the proposed method, three kinds of simulations with a spherical head model and cortical surface constraint were conducted. The measurements were simulated according to the configuration of a whole head MEG system (Vectorview 306, Neuromag Ltd. , Finland). The T1-weighted MR volume image used in this work was acquired from a normal subject on a 1.5 Tesla GE MR scanner by means of a three-dimensional sequence (TE = 1.828 ms, TR = 8.54 ms, flip angle = 15°, FOV = 26 × 26 × 10cm³, matrix size = 256 × 256, voxel size = 1.02 × 1.02 × 1.50mm³). The cortical surface was reconstructed by the software FreeSurfer [12] and consisted of 114,024 vertices. In addition to the

specified dipole sources described in the following sessions, each simulation also contained 3000 random dipoles which were evenly distributed within the sphere with radius of 7 cm and with standard deviation of 0.1 nAm.

Location error (LE) and similarity of temporal activities were used as the index of accuracy in this work. LE was estimated by calculating the distance between the location of the peak on the reconstructed cortical source distribution and the ground truth. Similarity between the temporal activity of the dipole sources and that of the selected components was calculated as their correlation coefficient.

3.1 Single Dipole Source

In the first simulation, the ground truth contained one single dipole, as shown in Fig. 1. It was a 15 Hz sine wave modulated with a Gaussian kernel (the blue dashed line in Fig. 1 (f)) and was placed at \mathbf{r}_1 with orientation \mathbf{q}_1 (the green point and green arrow in Fig. 1 (a)). Fig. 1 (c) displays the corresponding lead field vector with parameter $\theta_1 = \{\mathbf{r}_1, \mathbf{q}_1\}$. Fig. 1 (e) shows the simulated measurement, whose topography at 250 ms is displayed in Fig. 1 (b).

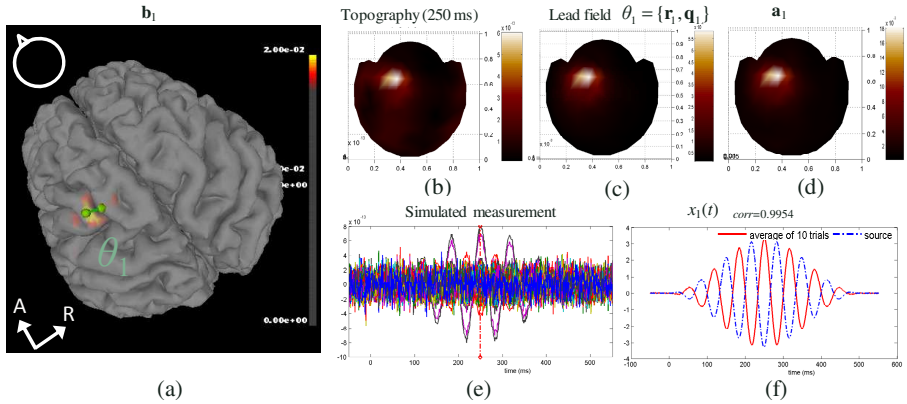


Fig. 1. Ground truth and results of simulation 1: (a) the location (green point) and orientation (green arrow) of the given dipole source, and the cortical source distribution of the interested IC (highlighted region), (b) the topography of simulated measurement at 250 ms, (c) the lead field vector with parameter $\theta_1 = \{\mathbf{r}_1, \mathbf{q}_1\}$, (d) the topography of the interested IC, which is extracted by second-stage ICA, (e) the simulated measurement, and (f) the comparison between the temporal profile of the given source (blue line) and that of the interested component (red line)

There were 105 ICs extracted from simulated measurements by the first-stage ICA, yet only one of these ICs was selected as a meaningful source and then used for the second-stage ICA. The estimated IC and the corresponding cortical source distribution are displayed in Figs. 1 (d) and (a). The LE and similarity is shown in Table 1.

Table 1. Location errors and similarities of the temporal profiles in the three simulations

Simulation	IC	Waveform	Frequency (Hz)	Similarity	LE (mm)
1	1	sine	15	-0.9954	4.67
2	1	tangent	11	0.9997	0.00
	2	sine	15	0.9953	0.91
3	1	sine	7	0.9886	3.64
	2	sine	17	-0.9851	3.62
	3	sine	31	-0.9810	0.00 ^a and 3.38 ^b

^a The distance between \mathbf{r}_{31} and the peak on the left hemisphere.

^b The distance between \mathbf{r}_{33} and the peak on the right hemisphere.

3.2 Two Uncorrelated Dipole Sources

Two uncorrelated dipole sources with parameters θ_{21} and θ_{22} were placed as shown in Fig. 2 (a). The temporal waveforms of these two sources were constructed from 11 Hz tangent and 15 Hz sine wave functions as shown in Fig. 2 (e), green and blue curves, respectively.

Two of the 93 ICs, extracted from simulated measurement by the first-stage ICA, were selected as the meaningful components and then reconstructed to be the de-noised measurement followed by the second-stage ICA. The extracted two interested ICs and the corresponding cortical source distributions are shown in Fig. 3. In this simulation, the spatiotemporal imaging of the two ICs were almost perfectly fit the ground truths with both of the LEs less than 1 mm (Table 1).

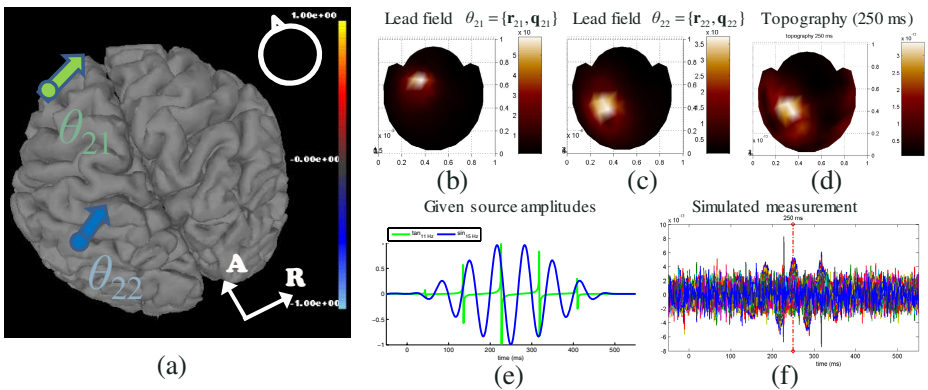


Fig. 2. Ground truth of simulation 2: (a) the locations and orientations of two simulated dipole sources with parameters θ_{21} (green) and θ_{22} (blue), (b) the lead field vector $\mathbf{l}_{\theta_{21}}$ for the first dipole, (c) the lead field vector $\mathbf{l}_{\theta_{22}}$ for the second dipole, (d) the topography of simulated measurement at 250 ms, (e) the given temporal profiles of two dipoles, and (f) the time courses of the simulation

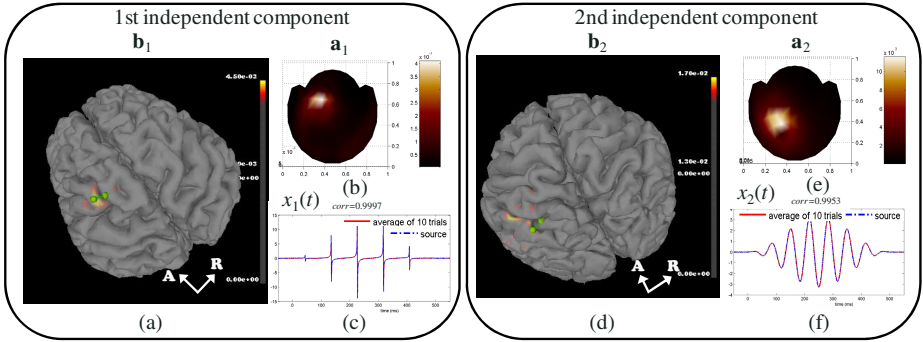


Fig. 3. The two interested independent components and the respective cortical source distributions in simulation 2

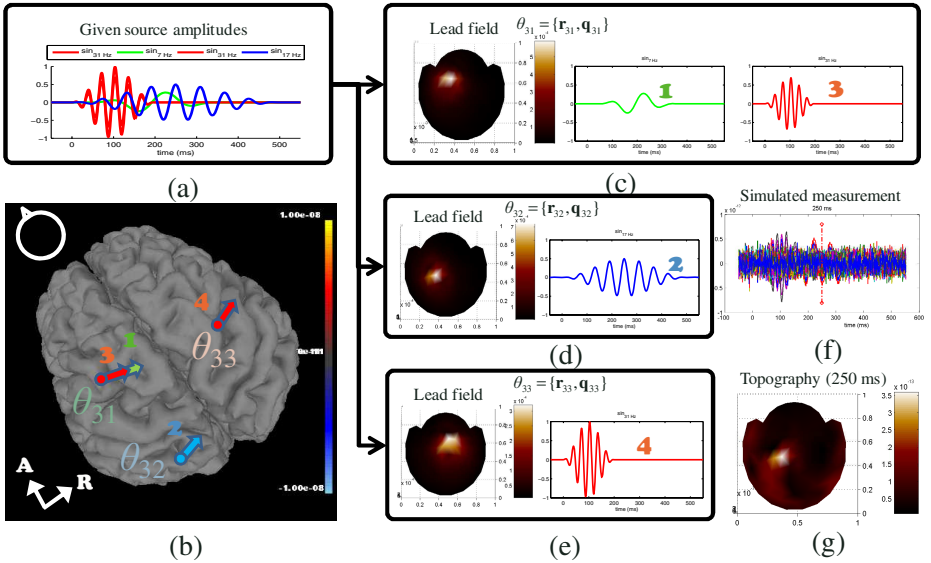


Fig. 4. Ground truth of simulation 3: (a) temporal waveforms of the four dipole sources (two of them have the same waveforms), (b) the locations and orientations of four dipoles labeled by different numbers and colors, (c) the lead field vector $\mathbf{l}_{\theta_{31}}$ and the temporal activities of two given sources numbered as 1 and 3, (d) the lead field vector $\mathbf{l}_{\theta_{32}}$ and the temporal activities of the second source, (e) the lead field vector $\mathbf{l}_{\theta_{33}}$ and the temporal activities of the fourth source, (f) the time courses of the simulated measurement, and (g) the topography of simulated measurement at 250 ms

3.3 Four Dipole Sources

In the third simulation, four dipole sources were placed at three distinct positions. Fig. 4 illustrates the temporal waveforms, locations, and orientations of

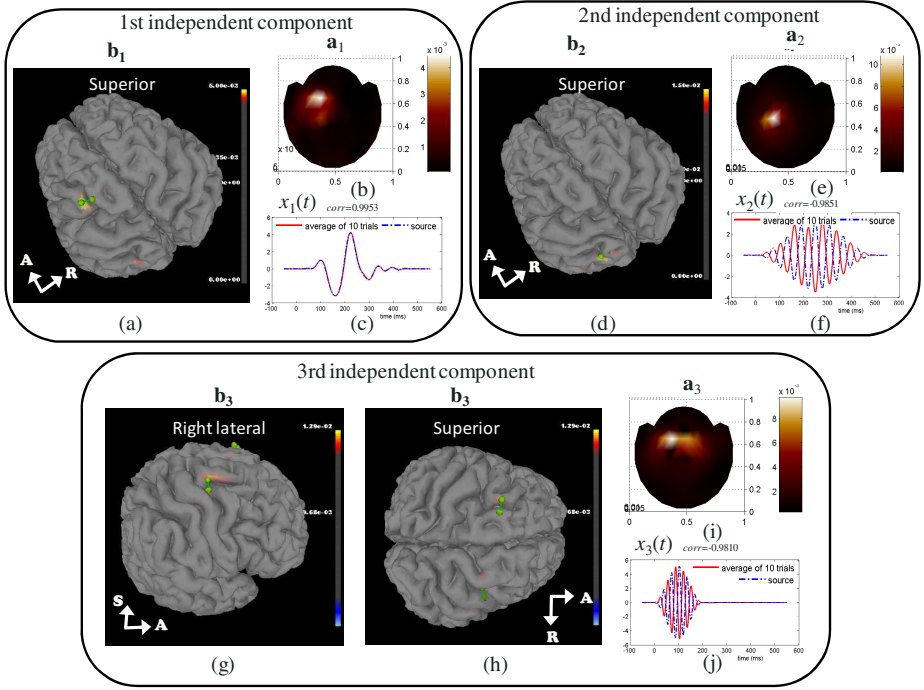


Fig. 5. The three interested independent components and the respective cortical source distributions in simulation 3

these dipoles. Two of the sources, the first and the third one, had the same location and orientation with the parameter $\theta_{31} = \{\mathbf{r}_{31}, \mathbf{q}_{31}\}$, but with different time waveforms, 7 Hz and 31 Hz sine waves, respectively.

During the first-stage ICA, 105 ICs were extracted from simulated measurement. Only three of the 105 components were chosen to reconstruct as the de-noised measurement. Finally, three interested ICs were calculated in the second-stage ICA and then used to map to the cortical surface as shown in Fig. 5. The estimated LEs and the similarities are shown in Table II.

According to the results of the three simulations, each simulation has at least one estimated source located at the position near the ground truth. The spike-like tangent wave in the second simulation has the highest similarity and the least location error (Table II). For the sources having the same waveform but placed at two different positions, that is, the third and the fourth dipoles with 31 Hz sine waves in simulation 3, the reconstructed cortical distribution is accurate.

4 Conclusions

We have presented a method of spatiotemporal source imaging for independent components extracted from conventional ICA algorithms. Compared to

EMSICA, the proposed method has the advantage of smaller amount of unknown parameters. In our experiments, the location error is within 5 mm for well-separated components. Besides, the proposed method can well map the cortical source distribution for independent components originating from different positions. Therefore, it might stand a chance for imaging distributed neural networks.

References

1. Lee, T., Girolami, M., Bell, A., Sejnowski, T.: A Unifying Information-Theoretic Framework for Independent Component Analysis. *Computers & Mathematics With Applications* 39(11), 1–21 (2000)
2. Jung, T.P., Makeig, S., Humphries, C., Lee, T.W., McKeown, M.J., Sejnowski, V.I.T.J.: Removing Electroencephalographic Artifacts by Blind Source Separation. *Psychophysiology* 37(2), 163–178 (2000)
3. Kawakatsu, M.: Application of ICA to MEG Noise Reduction. In: 4th International Symposium of Independent Component Analysis and Blind Signal Separation (ICA 2003), pp. 535–541. Tokyo Denki University, Chiba, Japan (2003)
4. Cao, J., Murata, N., Amari, S., Cichocki, A., Takeda, T.: A Robust Approach to Independent Component Analysis of Signals with High-Level Noise Measurements. *IEEE Transactions On Neural Networks* 14(3), 631–645 (2003)
5. Escudero, J., Hornero, R., Abasolo, D., Fernandez, A., Lopez-Coronado, M.: Artifact Removal in Magnetoencephalogram Background Activity with Independent Component Analysis. *IEEE Transactions On Biomedical Engineering* 54(11), 1965–1973 (2007)
6. Mantini, D., Franciotti, R., Romani, G.L., Pizzella, V.: Improving MEG Source Localizations: An Automated Method for Complete Artifact Removal Based on Independent Component Analysis. *NeuroImage* 40(1), 160–173 (2008)
7. Qin, L., Ding, L., He, B.: Motor Imagery Classification by Means of Source Analysis for Brain Computer Interface Applications. *Journal of Neural Engineering* 1(3), 135–141 (2004)
8. Breun, P., Grosse-Wentrup, M., Utschick, W., Buss, M.: Robust MEG Source Localization of Event Related Potentials: Identifying Relevant Sources by Non-Gaussianity. In: Franke, K., Müller, K.-R., Nickolay, B., Schäfer, R. (eds.) *DAGM 2006*. LNCS, vol. 4174, pp. 394–403. Springer, Heidelberg (2006)
9. Tsai, A.C., Liou, M., Jung, T.P., Onton, J.A., Cheng, P.E., Huang, C.C., Duann, J.R., Makeig, S.: Mapping Single-Trial EEG Records on the Cortical Surface through a Spatiotemporal Modality. *NeuroImage* 32(1), 195–207 (2006)
10. Mosher, J.C., Leahy, R.M., Lewis, P.S.: EEG and MEG: Forward Solutions for Inverse Methods. *IEEE Trans. Biomed. Eng.* 46(3), 245–259 (1999)
11. Baillet, S., Mosher, J., Leahy, R.: Electromagnetic Brain Mapping. *IEEE Signal Processing Magazine* 18(6), 14–30 (2001)
12. Dale, A.M., Fischl, B., Sereno, M.I.: Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. *NeuroImage* 9(2), 179–194 (1999)

Morphological Hetero-Associative Memories Applied to Restore True-Color Patterns

Roberto A. Vázquez and Humberto Sossa

Centro de Investigación en Computación – IPN
Av. Juan de Dios Batíz, esquina con Miguel Othón de Mendizábal
Ciudad de México, 07738, México
ravem@ipn.mx, hsossa@cic.ipn.mx

Abstract. Morphological associative memories (MAMs) are a special type of associative memory which exhibit optimal absolute storage capacity and one-step convergence. This associative model substitutes the additions and multiplications by additions/subtractions and maximums/minimums. This type of associative model has been applied to different pattern recognition problems including face localization and reconstruction of gray scale images. Despite of his power, it has not been applied to problems involving true-color patterns. In this paper we describe how a Morphological Hetero-associative Memory (MHAM) can be applied in problems that involve true-color patterns. In addition, a study of the behavior of this associative model in the reconstruction of true-color images is performed using a benchmark of 14400 images altered by different type of noises.

Keywords: Associative memory, True-color patterns.

1 Introduction

The concept of associative memory AM emerges from psychological theories of human and animals learning. These memories store information by learning correlations among different stimuli. When a stimulus is presented as a memory cue, the other is retrieval as a consequence; this means that the two stimuli have become associated each other in the memory.

An AM can be seen as a particular type of neural network designed to recall output patterns in terms of input patterns that can appear altered by some kind of noise. Several AMs have been proposed in the last years. Refer for example to [1], [2], [3], [4], [5], [6], [7], [8], [9], [10] and [11]). Most of these AMs have several constraints that limit their applicability in complex problems. Among these constraints we could mention their capacity of storage (limited), the type of patterns (only binary, bipolar, integer or real patterns), robustness to noise (additive, subtractive, mixed, Gaussian noise, deformations, etc).

A first attempt in formulating useful morphological neural networks was proposed by Davidson et al. [12]. Since then, only a few papers involving morphological neural networks have appeared. Refer for example to [13] and [14]. In 1998, Ritter et al. [8] proposed the concept of morphological associative memories (MAMs). Basically, the

authors substituted the outer product by **max** and **min** operations. One year later, the authors introduced their morphological bidirectional associative memories [15]. Their properties, compared with Hopfield Associative model are completely different. For example, they exhibit optimal absolute storage capacity and one-step convergence in the auto-associative case. The MAM has been applied to different pattern recognition problems including face localization and reconstruction of gray scale images [9] and [16-20]. Despite of his power, it has not been applied in problems that involve true-color patterns neither a deep study of the Morphological Hetero-Associative Memory MHAM under true-color image patterns.

In this paper, it is described how a MHAM can be applied in problems that involve true-color patterns. Furthermore, a study of the behavior of the MHAM in the reconstruction of true-color images is performed using a benchmark of 14400 images altered by different type of noises.

2 Basics on Morphological Associative Memories

The basic computations occurring in the morphological network proposed by Ritter et al. are based on the algebraic lattice structure $(R, \wedge, \vee, +)$ where the symbols \wedge and \vee denote the binary operations of minimum and maximum, respectively.

Let $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$ an input and output pattern, respectively. An association between input pattern \mathbf{x} and output pattern \mathbf{y} is denoted as $(\mathbf{x}^\xi, \mathbf{y}^\xi)$, where ξ is the corresponding association. Associative memory \mathbf{W} is represented by a matrix whose components w_{ij} can be seen as the synapses of the neural network. If $\mathbf{x}^\xi = \mathbf{y}^\xi \forall \xi = 1, \dots, p$ then \mathbf{W} is auto-associative, otherwise it is hetero-associative. A distorted version of a pattern \mathbf{x} to be recuperated will be denoted as $\tilde{\mathbf{x}}$. If an AM \mathbf{W} is fed with a distorted version of \mathbf{x}^ξ and the output obtained is exactly \mathbf{y}^k , we say that recalling is robust.

Suppose we are given a couple of $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$. A morphological associative memory that will recall the pattern \mathbf{y} when presented the pattern \mathbf{x} is given by:

$$\mathbf{W} = \mathbf{y} \boxminus (-\mathbf{x})^t = \begin{pmatrix} y_1 - x_1 & \cdots & y_1 - x_n \\ \vdots & \ddots & \vdots \\ y_m - x_1 & \cdots & y_m - x_n \end{pmatrix} \quad (1)$$

since \mathbf{W} satisfies the equation $\mathbf{W} \boxminus \mathbf{x} = \mathbf{y}$ as can be verified by the simple computation

$$\mathbf{W} \boxtimes \mathbf{x} = \begin{pmatrix} \bigvee_{i=1}^n (y_1 - x_i + x_i) \\ \vdots \\ \bigvee_{i=1}^n (y_m - x_i + x_i) \end{pmatrix} = \mathbf{y} \tag{2}$$

\mathbf{W} is called the max product of \mathbf{y} and \mathbf{x} . We can also denote the min product of \mathbf{y} and \mathbf{x} using operator \boxdot .

For a given set of pattern associations $\{(\mathbf{x}^\xi, \mathbf{y}^\xi) : \xi = 1, \dots, k\}$ a couple of pattern matrices (\mathbf{X}, \mathbf{Y}) is defined, where $\mathbf{X} = (\mathbf{x}^1, \dots, \mathbf{x}^k)$, $\mathbf{Y} = (\mathbf{y}^1, \dots, \mathbf{y}^k)$. With each pair of matrices (\mathbf{X}, \mathbf{Y}) , two natural morphological $m \times n$ memories $\mathbf{W}_{\mathbf{XY}}$ and $\mathbf{M}_{\mathbf{XY}}$ are defined by:

$$\mathbf{W}_{\mathbf{XY}} = \bigwedge_{\xi=1}^k \left[\mathbf{y}^\xi \boxtimes (-\mathbf{x}^\xi) \right] \tag{3}$$

and

$$\mathbf{M}_{\mathbf{XY}} = \bigvee_{\xi=1}^k \left[\mathbf{y}^\xi \boxdot (-\mathbf{x}^\xi) \right]. \tag{4}$$

From this definition it follows that

$$\mathbf{y}^\xi \boxtimes (-\mathbf{x}^\xi)^t = \mathbf{y}^\xi \boxdot (-\mathbf{x}^\xi)^t \tag{5}$$

which implies that $\forall \xi = 1, \dots, k$

$$\mathbf{W}_{\mathbf{XY}} \leq \mathbf{y}^\xi \boxtimes (-\mathbf{x}^\xi)^t = \mathbf{y}^\xi \boxdot (-\mathbf{x}^\xi)^t \leq \mathbf{M}_{\mathbf{XY}}. \tag{6}$$

In terms of equations 2, 3 and 4, this last set of inequalities implies that $\forall \xi = 1, \dots, k$

$$\mathbf{W}_{\mathbf{XY}} \boxtimes \mathbf{x}^\xi \leq \left[\mathbf{y}^\xi \boxtimes (-\mathbf{x}^\xi)^t \right] \boxtimes \mathbf{x}^\xi = \mathbf{y}^\xi = \left[\mathbf{y}^\xi \boxdot (-\mathbf{x}^\xi)^t \right] \boxdot \mathbf{x}^\xi \leq \mathbf{M}_{\mathbf{XY}} \boxdot \mathbf{x} \tag{7}$$

or equivalently, that

$$\mathbf{W}_{\mathbf{XY}} \boxtimes \mathbf{X} \leq \mathbf{Y} \leq \mathbf{M}_{\mathbf{XY}} \boxdot \mathbf{X} \tag{8}$$

The complete set of theorems which guarantee perfect recall and their corresponding proofs are presented in [8]. Some important to mention is that this MAM is robust

to additive noise or subtractive noise, not both (mixed noise). While MAM \mathbf{W}_{XY} is robust to subtractive noise, MAM \mathbf{M}_{XY} is robust to additive noise.

3 Behavior of \mathbf{W}_{XY} under True-Color Patterns

In this section a study of the behavior of MHAM \mathbf{W}_{XY} under true-color noisy patterns is presented. First to all, we verified if the MHAM was capable to recall the complete set of associations. Then we verified the behavior of MHAM using noisy versions of the images used to train the MHAM. After that, we performed a study of how the number of associations influenced the behavior of the MHAM.

















0 % of noise				
10 % of noise				
30 % of noise				
50 % of noise				
	Additive noise	Subtractive noise	Mixed noise	Gaussian noise

Fig. 1. Some images which compose the benchmark used to train and test the MAM

The benchmark used in this set of experiments is composed by 14400 color images of 63×43 pixels and 24 bits in a bmp format. This benchmark contains 40 classes of flowers and animals. Per each class, there are 90 images altered with additive noise (0% of the pixels to 90% of the pixels), 90 images altered with subtractive noise (0% of the pixels to 90% of the pixels), 90 images altered with mixed noise (0% of the pixels to 90% of the pixels) and 90 images altered with Gaussian noise (0% of the pixels to 90% of the pixels). In Fig. 1 some images which compose this benchmark are shown.

Before the MHAM \mathbf{W}_{XY} was trained, each image was transformed into an image pattern. To build an image pattern from the bmp file, the image was read from left-right and up-down; each RGB pixel (hexadecimal value) was transformed into a decimal value and finally, this information was stored into an array. Once trained the associative memory, we proceeded to evaluate the behavior of the MHAM. In order to measure the accuracy of the MHAM we counted the number of pixels correctly recalled.

For the first set of experiments, we found that MHAM \mathbf{W}_{XY} was not capable to recall the complete set of associations. Even when patterns were not altered with

noise, the MHAM correctly recalled only the 77.6% of the pixels. Studying the behavior of the MHAM under different type of noises and trained with 20 associations we found that, for the case of additive noise, if only the 2 % of the pixels are altered, the MHAM was capable of correctly recall only the 2.2% of the pixels. For the case of mixed and Gaussian noise, we observed that if only the 2 % of the pixels are altered, the MHAM is capable of correctly recall only the 3.4% and 17.7 % of the pixels, respectively. These percentages decreased when the number of altered pixels was increased. For the case of subtractive noise, we observed that even when the 90% of the pixels are altered, the MHAM was capable of correctly recall the 73.8% of the pixels. In average, for the case of the image patterns altered with additive noise the MHAM recalled the 1% of the pixels. For the case of subtractive noise the MHAM in average recalled the 75.9% of the pixels. For the case of mixed and Gaussian noise, the MHAM recalled the 1.19 % and 2.7% respectively.

By analyzing the behavior of the MHAM \mathbf{W}_{XY} under different type of noises and trained with 10 associations we found that when patterns were not altered with noise, the MHAM correctly recalled only the 90% of the pixels. For the case of additive, mixed and Gaussian noise, we observed that if only the 2 % of the pixels were altered with additive noise, the MHAM was capable of correctly recalling only the 5.4%, 5.5% and 14.1% of the pixels, respectively. For the case of subtractive noise, we observed that even when the 90% of the pixels were altered, the MHAM was capable of correctly recalling the 87.8% of the pixels. In average, for the case of the image patterns altered with additive noise the MHAM recalled the 1.4% of the pixels. For the case of subtractive noise the MHAM in average recalled the 88.9% of the pixels. For the case of mixed and Gaussian noise, the MHAM recalled the 1.6 % and 4.1% respectively.

Finally, we studied the behavior of the MHAM \mathbf{W}_{XY} under different type of noises and trained with 5 associations. When patterns were not altered with noise, the MHAM correctly recall only the 98.8% of the pixels. For the case of additive, mixed and Gaussian noise, we observed that if only the 2 % of the pixels are altered with additive noise, the MHAM was capable of correctly recalling only the 6.27%, 3.14% and 17.2% of the pixels, respectively. For the case of subtractive noise, we observed that even when the 90% of the pixels are altered, the MHAM was capable of correctly recalling the 98.8% of the pixels. In average, for the case of the image patterns altered with additive noise the MHAM recalled the 1.4% of the pixels. For the case of subtractive noise the MHAM in average recalled the 98.4% of the pixels. For the case of mixed and Gaussian noise, the MHAM recalled the 1.5 % and 4.2% respectively.

In short, we can say that the accuracy of the MHAM increased as the number of associations is decreased. This fact holds only for patterns altered with subtractive noise. For the other type of noises tested in this set of experiments, the accuracy decreased.

The general behavior of the MHAM \mathbf{W}_{XY} is shown in Fig. 2, where clearly we can observe the robustness of this memory with patterns altered with subtractive noise.

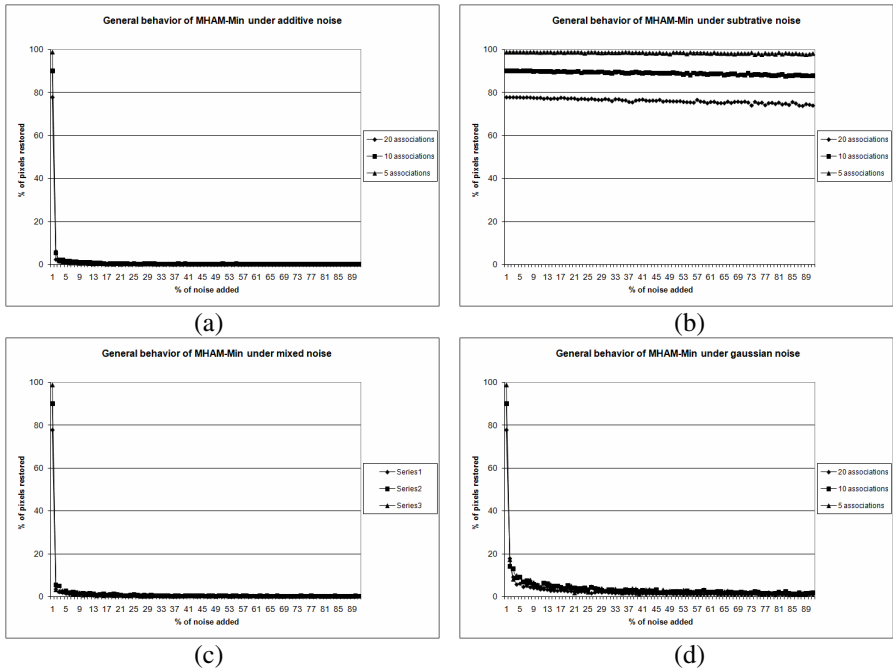


Fig. 2. General behavior of the MHAM W_{XY} under different type of noises

4 Behavior of M_{XY} under True-Color Patterns

In this section a study of the behavior of MHAM M_{XY} under true-color noisy patterns is presented. First to all, we verified if the MHAM was capable to recall the complete set of associations. Then we verified the behavior of MHAM using noisy versions of the images used to train the MHAM. After that, we performed a study of how the number of associations influenced the behavior of the MHAM.

The benchmark used in this set of experiments was the same used in section 3. Before each MHAM M_{XY} was trained, each image was transformed into an image pattern. Once trained the associative memory, we proceed to evaluate the behavior of the MHAM.

Through this set of experiments, we realized that MHAM M_{XY} was not capable to recall the complete set of associations. Even when the patterns were not altered with noise, the MHAM correctly recalled only the 43.1% of the pixels. By analyzing the behavior of the MHAM under different type of noises and trained with 20 associations we found that, for the case of subtractive, mixed and Gaussian noise, if only the 2% of the pixels were altered, the MHAM was capable of correctly recalling only the 7.2%, 10% and 22.6% of the pixels, respectively. These percentages decreased when the number of altered pixels was increased. For the case of additive noise, we observed that even when

the 90% of the pixels are altered, the MHAM was capable of correctly recall the 39.5% of the pixels. In average, for the case of the image patterns altered with additive noise the MHAM recalled the 41.3% of the pixels. For the case of subtractive noise the MHAM recalled the 1.2% of the pixels. For the case of mixed and Gaussian noise, the MHAM recalled the 1.8 % and 6.7% respectively.

By analyzing the behavior of the MHAM M_{XY} under different type of noises and trained with 10 associations we found that when patterns were not altered with noise, the MHAM correctly recalled only the 67.3% of the pixels. For the case of subtractive, mixed and Gaussian noise, we observed that if only the 2 % of the pixels are altered, the MHAM was capable of correctly recalling only the 11.5%, 17.3% and 37% of the pixels, respectively. For the case of additive noise, we observed that even when the 90% of the pixels are altered, the MHAM was capable of correctly recalling the 61.3% of the pixels. In average, for the case of the image patterns altered with additive noise the MHAM recalled the 65% of the pixels. For the case of subtractive noise the MHAM recalled the 1.7% of the pixels. For the case of mixed and Gaussian noise, the MHAM recalled the 2.6 % and 9.6% respectively.

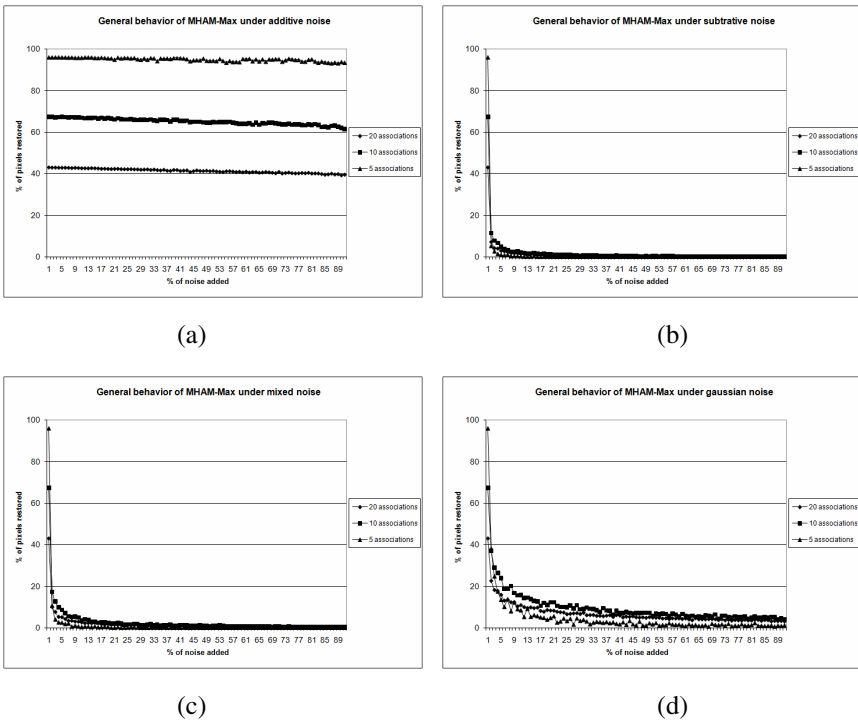


Fig. 3. General behavior of the MHAM under different type of noises

Finally, we studied the behavior of the MHAM \mathbf{M}_{XY} under different type of noises and trained with 5 associations. When patterns were not altered with noise, the MHAM correctly recalled only the 96.1% of the pixels. For the case of subtractive, mixed and Gaussian noise, we observed that if only the 2 % of the pixels are altered, the MHAM was capable of correctly recalling only the 5.3%, 10.7% and 38.1% of the pixels, respectively. For the case of additive noise, we observed that even when the 90% of the pixels were altered, the MHAM was capable of correctly recalling the 93.5% of the pixels. In average, for the case of the image patterns altered with additive noise the MHAM recalled the 95% of the pixels. For the case of subtractive noise the MHAM in average recalled the 1.2% of the pixels. For the case of mixed and Gaussian noise, the MHAM recalled the 1.4 % and 4.8% respectively.

In short, we can say that the accuracy of the MHAM \mathbf{M}_{XY} increased as the number of associations decreased. This fact holds only for patterns altered with additive noise. For the other type of noises tested in this set of experiments, the accuracy decreased.

The general behavior of the MHAM \mathbf{M}_{XY} is shown in Fig. 3, where clearly we can observe the robustness of this memory with patterns altered with additive noise.

5 Conclusions

In this paper, a complete study of the behavior of the morphological hetero-associative memory in the restoration of true-color images was performed using a benchmark of 14400 images altered by different type of noises.

Due to this associative model had been only applied to binary and gray level patterns, this paper is useful to better understand the power and limitations of this model. Two types of experiments were performed. In the first case we studied the hetero-associative version of MHAM \mathbf{W}_{XY} . In the second case we studied the hetero-associative version \mathbf{M}_{XY} . For both cases we verified if the MAM was capable to recall the complete set of associations. Then we verified the behavior of the MAM using noisy versions of the images used to build the memory. After that, we performed a study of how the number of associations influences the behavior of the MAM.

Through several experiments, we found some interesting properties of this associative model. MHAMs do not present perfect recall but are not sensitive to the amount of noises. In other words, MHAMs hold the accuracy even when the noise was increased.

As we already knew, MHAM \mathbf{M}_{XY} is more robust to additive noises than the other type of noises. However, MHAM \mathbf{M}_{XY} is more robust to Gaussian and mixed noise than subtractive noise. For the case of MHAM \mathbf{W}_{XY} , this memory is more robust to subtractive noises than the other type of noises. However, MHAM \mathbf{W}_{XY} is more robust to Gaussian and mixed noise than subtractive noise.

Regarding to the storage capacity, we found that the accuracy of the model is too sensitive to the number of associations stored in the MAM. In general we can say that when the number of association is increased, the accuracy of the memory decreases when patterns are altered with noise to which they are more robust. If patterns are altered with noise to which they are not robust, the accuracy of the memory increases when the number of associations is increased.

On the other hand, the accuracy of the hetero-associative memory decreases when the amount of additive noise is increased. The best accuracy is provided by \mathbf{W}_{XY} when patterns are altered by subtractive noise. In average, MHAM \mathbf{M}_{XY} correctly recall 67.1% of the pixels when patters are altered by additive noise. MHAM \mathbf{W}_{XY} correctly recall 87.7% of the pixels when patterns are altered by additive noise.

Acknowledgments. This work was economically supported by SIP-IPN under grant 20082948 and CONACYT under grant 46805.

References

1. Steinbuch, K.: Die Lernmatrix. *Kybernetik* 1, 26–45 (1961)
2. Anderson, J.A.: A Simple Neural Network Generating an Interactive Memory. *Math. Biosci.* 14, 197–220 (1972)
3. Kohonen, T.: Correlation matrix memories. *IEEE Trans. on Comp.* 21, 353–359 (1972)
4. Hopfield, J.J.: Neural Networks and Physical Systems with Emergent Collective Computational Abilities. *Proc. Natl. Acad. Sci.* 79, 2554–2558 (1982)
5. Sussner, P.: Generalizing Operations of Binary Auto-associative Morphological Memories using Fuzzy Set Theory. *J. Math. Imaging Vis.* 19, 81–93 (2003)
6. Ritter, G.X., et al.: Reconstruction of Patterns from Noisy Inputs using Morphological Associative Memories. *J. Math. Imaging Vis.* 19, 95–111 (2003)
7. Sossa, H., Barrón, R., Vázquez, R.A.: Transforming Fundamental Set of Patterns to a Canonical Form to Improve Pattern Recall. In: Lemaître, C., Reyes, C.A., González, J.A. (eds.) *IBERAMIA 2004. LNCS*, vol. 3315, pp. 687–696. Springer, Heidelberg (2004)
8. Ritter, G.X., Sussner, P., Diaz de Leon, J.L.: Morphological Associative Memories. *IEEE Trans Neural Networks* 9, 281–293 (1998)
9. Sussner, P., Valle, M.: Gray-Scale Morphological Associative Memories. *IEEE Trans. on Neural Netw.* 17, 559–570 (2006)
10. Vazquez, R.A., Sossa, H.: A New Associative Memory with Dynamical Synapses. *Neural Processing Letters* 28, 189–207 (2008)
11. Vazquez, R.A., Sossa, H.: A Bidirectional Heteroassociative Memory for True Color Patterns. *Neural Processing Letters* 28, 131–153 (2008)
12. Davidson, J.L., Ritter, G.X.: A Theory of Morphological Neural Networks. In: *Digital Optical Computing*, pp. 378–388. SPIE, Los Angeles (1990)
13. Suarez-Araujo, C.P.: Novel Neural Network Models for Computing Homothetic in Variances: An Image Algebra Notation. *J. Math. Imaging and Vision* 7, 69–83 (1997)
14. Davidson, J.L., Hummer, F.: Morphology Neural Networks: An Introduction with Applications. *IEEE System Signal Processing* 12, 177–210 (1993)
15. Ritter, G.X., Diaz-de-Leon, J.L., Sussner, P.: Morphological Bidirectional Associative Memories. *Neural Networks* 12, 851–867 (1999)

16. Raducanu, B., Graña, M., Albizuri, X.F.: Morphological Scale Spaces and Associative Morphological Memories: Results on Robustness and Practical Applications. *J. Math. Imaging and Vision* 19, 113–131 (2003)
17. Graña, M., Gallego, J., Torrealdea, F.J., D’Anjou, A.: On the Application of Associative Morphological Memories to Hyperspectral Image Analysis. In: Mira, J., Álvarez, J.R. (eds.) *IWANN 2003*. LNCS, vol. 2687, pp. 567–574. Springer, Heidelberg (2003)
18. Sussner, P.: Associative Morphological Memories Based on Variations of the Kernel and Dual Kernel Methods. *Neural Netw.* 16, 625–632 (2003)
19. Wang, M., Wang, S.T., Wu, X.J.: Initial Results on Fuzzy Morphological Associative Memories. *ACTA ELECTRONICA SINICA (in Chinese)* 31, 690–693 (2003)
20. Feng, N., Qiu, Y., Wang, F., Sun, Y.: A Unified Framework of Morphological Associative Memories. In: *ICIC 2006*, pp. 1–11. Springer, Berlin (2006)

A Novel Method for Analyzing Dynamic Complexity of EEG Signals Using Symbolic Entropy Measurement

Lisha Sun¹, Jun Yu¹, and Patch J. Beadle²

¹ The IMT Key Lab of State Education Ministry, College of Engineering,
Shantou University, Guangdong, 515063, China
lssun@stu.edu.cn

² School of System Engineering, The University of Portsmouth, Portsmouth, U.K.

Abstract. Symbolic entropy is proposed to measure the complexity of the electroencephalogram (EEG) signal under different brain functional states. The EEG data recorded from different subjects were investigated and compared with both approximate entropy (ApEn) and Shannon entropy. The experimental results show that the proposed method can effectively distinguish the complexities of two groups. The experimental results provide preliminary support for the notion that the complex nonlinear nature of brain electrical activity may be the result of isolation or impairment of the neural information transmission within the brain. It is concluded that symbolic entropy serves a better measure for EEG signals and other medical signals.

Keywords: Approximate Entropy, Shannon Entropy, EEG.

1 Introduce Time Series Analysis

Many evidences have been found that human brain is a complicated nonlinear spatial-temporal neural system [1,2]. The highly complex nonlinear system of human brain shows the chaotic dynamics. Advanced approaches for studying EEG signal enable us to extract more useful information and understand the underlying inherent mechanism under different brain functional states. Nonlinear dynamics theory brought us some new sights. It is important in describing a large number of complex physiological systems and complex time series such as EEG, using adequate nonlinear dynamical analysis rather than linear time series analysis [3]. In principle, nonlinear dynamics can provide a more complete description of the EEG recordings and a better understanding of the underlying mechanism of brain, such as Lyapunov exponent and correlation dimensions [4-6]. However, these techniques were based on the low-dimension nonlinear dynamics system. Many algorithms from nonlinear dynamics and theory of deterministic chaos were found chronically unreliable, often producing spurious dimension or Lyapunov exponent estimates and thus supporting false identification of chaotic dynamics existing in the observed data [7].

Based on the common information theory, the quantification of complexity measurement is used as the nonlinear detection of EEG time series is. Several entropies for identifying the complexity of medical signals have been introduced. Since the

pioneer work of Pincus, the approximate entropy (ApEn) has been widely used in measuring the regularity or complexity of biomedical signals [8-11]. Moreover, several entropies based on symbolic dynamics for identifying the complexity of medical signals were also discussed and obtained a rapidly development, making it an essential part of the nonlinear data analysis such as weather forecast, neural network, and information processing, especially in biological systems [12-18]. In [12], ApEn was used to analyze the irregularities and nonlinearities in fetal heart period time series in the course of pregnancy. Another entropy approach employed for quantifying the regularity of a time series was Shannon entropy. The entropy of short binary sequences in heart period dynamics was also examined in [13]. Novel symbolic entropy was proposed recently [19,20]. The goal of this paper is to extend the dynamic entropy approach to the EEG analysis. By setting the adequate threshold, the nonlinear dynamic time series are converted into symbolic system and a new kind of complexity analysis in terms of symbolic entropy is presented. Both simulated data and real EEG data are examined using the proposed algorithm and compared with the traditional entropy approaches.

The paper is organized as follows: In section 2, both binary Shannon entropy and symbolic entropy are introduced. The symbolic entropy is applied to test the logistic map and the one-way coupled map lattice in section 3. Some useful results are also provided. In section 4, the symbolic entropy is used to deal with the EEG data and compare the traditional entropies with the presented method. The results and discussions are also given.

2 Symbolic Entropy

We consider symbolic dynamics as a coarse grained description of trajectories of a general class of systems, which remains both robust and statistical properties of the system invariant. The basic principle of symbolic dynamics is to transform a time series into a symbol sequence, which provides a model for the orbits of the dynamical system via a space of sequences.

For a given data set X , the symbol sequence is achieved by quantifying X into boxes labeled with a symbol. Calculating the attributes of the symbol sequence can reveal the nonlinear characteristics of the original time series and the underlying dynamical system. If symbolic dynamics is described in the right way, the ensemble of trajectories in a trajectory set has its common statistical properties.

Let $\{H_i \mid i = 0, 1, 2, \dots, L-1\}$ be a family of L disjunctive subsets covering the whole state space X , i.e. $\bigcap_{i=0}^{L-1} H_i = \Phi$, $\bigcup_{i=0}^{L-1} H_i = X$. The index set $H = \{0, 1, 2, \dots, L-1\}$ of the partition can be interpreted as a finite alphabet of letters: $H_i = i$.

Symbolic entropy is based on quantity binary Shannon entropy (BinShan) [13]. One of the differences is that BinShan is based on the binary sequences, while the symbolic entropy is based on multiple integers. Another differences is that the partition rules.

Considering a given symbol set composed of $\{s_0, s_1, \dots, s_{K-1}\}$ and a data set composed of $K+1$ critical points $\{c_0, c_1, \dots, c_K\}$. The given chaotic sequences

$\{x_i \mid i = 1, 2, \dots, N\}$ can be replaced by the symbol sequences $\{s_i \mid i = 1, 2, \dots, N\}$, using the following partition rule [19, 20]:

$$\text{if } c_k < x_i \leq c_{k+1}, \text{ then } s_i = s_k \tag{1}$$

The time series is converted into pseudo-random sequences such that

$$s_i(x) = j, \text{ if } \sin^2 \left| \frac{j\pi}{2K} \right| < x_i \leq \sin^2 \left| \frac{(j+1)\pi}{2K} \right| \quad j = 0, 1, 2, \dots, K-1 \tag{2}$$

K is the total number of different symbols. Then, the pseudo-random sequences $\{s_i \mid i = 1, 2, \dots, N\}$ are segmented into a set of short sequences with length of L . Pseudo-random sequences of vectors $u(1), \dots, u(N - L + 1)$ are formed by defining $u(i) = [s(i), s(i + 1), \dots, s(i + L - 1)]$. Consequently, it can be marked and identified as

$$l_x(L, i) = \sum_{p=1}^L K^{L-p} s(p + i) \tag{3}$$

where K denotes the number of different integers in $\{s_i \mid i = 1, 2, \dots, N\}$ and L is the length of the short sequences. i represents the beginning point of the symbol set. By quantifying the time series derived from the practical system, the symbol s_k can be replaced with an integer K . The pseudo-random sequences can be easily identified with the data set $\{0, 1, \dots, K^L - 1\}$. Thus the symbolic entropy of the pseudo-random sequences can be defined to quantify the information involved in the symbolic sequences:

$$E = -\frac{1}{L} \sum_{l_x} P_{l_x} \ln P_{l_x} \tag{4}$$

where $P_{l_x} = \frac{n_{l_x}}{n_{sum}}$ denotes the probability of the pattern l_x . n_{l_x} is the number of occurrences of the pattern l_x , whereas n_{sum} represents the total number of the patterns.

Obviously, information contained in the symbolic entropy is related with the number of the critical points. If the number of c_i is given, we can find the best critical values by optimizing the symbolic entropy E . The number of the critical points is proportional to the ability of coding the original time series. Furthermore, the more the critical points, the higher the symbolic entropy. Thus only few distinct patterns occur for a very regular binary sequence. The symbolic entropy is very small since the

probability of these patterns is very high and only little information is contained in the whole sequence.

3 The Complexity of Chaotic Series

The behaviors of symbolic entropy with two chaotic sequences is evaluated and compared. Firstly, the complexity of simple chaotic sequences is investigated. Consider the well-known logistic map, which is defined as follows:

$$x(n+1) = rx(n)[1-x(n)] \quad (5)$$

where $r \in [0,4]$, $x \in [0,1]$, $n \geq 0$.

The complexity of logistic map depends on the control parameter r , which displays chaotic property when $r \in [3.5,4.0]$. Fig. 1 depicts the Lyapunov exponent λ within this region. Obviously, the maximum is obtained at $r = 4.0$. When r lies in $[3.80, 3.85]$, λ decreases sharply and reaches to the minimum.

Lyapunov exponent gives a better description of the system complexity. If the characterized complexity measure is similar to this criterion, the algorithm is available. In this study, we focus on the entropy-based complexity. Fig. 2 provides the result of the symbolic entropy and ApEn for the logistic map with the same r as shown in Fig. 1, respectively. It is noticed that the curves of entropies are much smoother than the λ . The reason is that the computation of entropies is based on statistics. Hence, it is a coarse description of system dynamics.

Generally speaking, both of two entropies show similar complexity, in accordance with the result from the Lyapunov exponent. However, for some special cases, such as in the interval $[3.80, 3.85]$ and $r = 3.5$, ApEn cannot provide a better result. From Fig.1, we can see that the Lyapunov exponent of the former case is distinctly smaller than the latter one, but the ApEn of these two cases keep the same. On the contrary, the symbolic entropy can effectively differentiate these weak differences with less computation complexity.

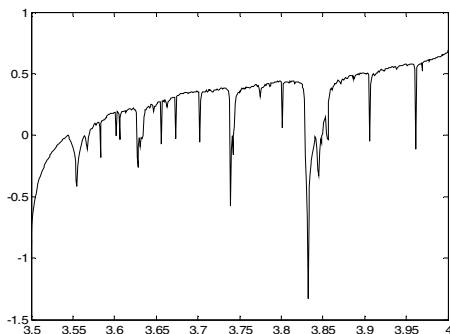


Fig. 1. The Lyapunov exponent for the logistic map

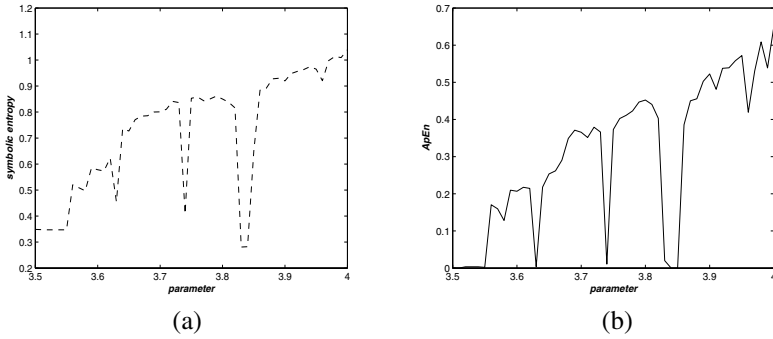


Fig. 2. The symbolic entropy (a) and ApEn (b) for the logistic map

As the second example, a pseudo-random sequence $\{g_n\}$ derived from one-way coupled map lattice is generated as:

$$x_{n+1}(1) = (1 - \varepsilon)f[0x_n(1)] + \varepsilon g_n \tag{6}$$

$$x_{n+1}(i) = (1 - \varepsilon)f[x_n(i)] + \varepsilon f[x_n(i + 1)] \tag{7}$$

where $i=2, \dots, 6; n=0, 1, 2, \dots, N$

$$g_n = f[x_n(2)] \tag{8}$$

where n denotes the discrete-time step and i stands for the lattice point in the logistic map: $f(x) = rx(1 - x)$ in which both r and ε are the parameters of the system.

Table.1 gives the experimental results with critical points $K=2^3$ and $K=2^4$, respectively. The results show that symbolic entropy can efficiently measure the complexity of chaotic sequences when $L \geq 2$. However, for $K=2^3$ or $K=2^4$, only eight or sixteen different short sequences were found for $L = 1$. For this case, the symbolic sequences have the uniform distribution and provide poor performance. The results illustrate that the pseudo-random sequence generated from one-way coupled map lattice is more complex than that produced from the logistic map when $L \geq 2$, which accords with the discussion above. With the increase of the critical points K , the increase of the symbolic entropy was homologous as shown in Table. 1. To compare the behaviors, we also estimate the BinShan and ApEn for the two sequences, as shown in Table.2 and Table.3 from which it can be seen that the BinShan is consistent with the symbolic entropy, but the difference between these two time series is not significant. The BinShan of the sequences from one-way coupled map lattice is larger than that from the logistic map. Table.3 indicates that the more complicated the chaotic series, the larger the ApEn. However, when the embedding dimensions get larger such as $m > 4$, the results is not reliable since the proposed value by Pincus is $m = 2$. In conclusion, all the three approaches can evaluate the complexity of the time series, but both BinShan and ApEn are significantly inferior to the symbolic entropy.

Table 1. The symbolic entropies of different time series for different K

L	K=2 ³		K=2 ⁴	
	Logistic map	g _n	Logistic map	g _n
1	2.0793	2.0523	2.7720	2.7328
2	1.3860	2.0492	1.7322	2.7197
3	1.1548	2.0415	1.3854	2.6534
4	1.0391	1.9931	1.2117	2.2705
5	0.9693	1.7919	1.1066	1.8402

Table 2. The BinShan for different time series with L=1, 2, 3, 4

L	L=1	L=2	L=3	L=4
Logistic map	0.9898	0.9898	0.9897	0.9429
g _n	0.9997	0.9997	0.9993	0.9517

Table 3. The ApEn for different time series with L=5, K=1, 2, 3, 4, 5

N=5	K=1	K=2	K=3	K=4	K=5
Logistic map	0.6541	0.6539	0.6500	0.6430	0.6365
g _n	1.9939	1.9919	1.8125	0.9872	0.3620

4 EEG Signal and Analysis

4.1 Complexity Analysis of EEG

It is important to consider the effect of coarse-grained processing according to the property of the real system [21,22]. Fig.3 demonstrates the simulation result with different numbers of the critical points for $K = 2, 4, 8, 16$ for a segment of the EEG data. When $K = 2, 4$, the symbolic entropy fluctuates at some points while the symbolic entropy shows more reliable with similar changing trend when $K \geq 8$. In addition, by increasing the numbers of the critical points, the computation complexity will obviously increase. Therefore, selection of $K = 8$ is recommended and used in the following discussions.

Based on the analysis above, the proposed symbolic entropy shows the good behaviours among these three algorithms and the results are heuristic. It shows that complexities of EEG data from the healthy group are similar. They are concentrated on a certain value with a smaller deviation by comparing the symbolic entropy of the schizophrenic EEG recordings.

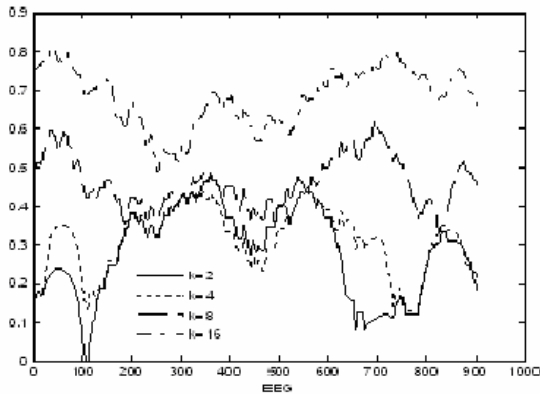


Fig. 3. The symbolic entropy with different critical points K

4.2 Results and Discussion

From the point of view with information theory, lower entropies indicate greater signal regularity corresponding to the situations in which communication pathways in a network are poorly developed or the system components operate in relative isolation. In contrast, entropies typically increase with greater system coupling and information exchanges among the systems [22]. The results of lower entropy value of EEG signals collected from the schizophrenic group may indicate a lower level of communication or information transmission among the different regions of the brain. Fig. 5 demonstrates a temporal-spatial plot with symbolic entropy of the EEG on the cerebral scalp. It is clear that the symbolic entropies lies in the left brain decreased are larger than that in the right due to the disorder of the brain. This experimental result is in accordance with the previous conclusion. Accordingly, it was hypothesized that the neural information transmission or communication for the schizophrenic patients between the main focus and other areas of the brain may be partly isolated or impaired. The low ability of communication would result in greater regularity of the brain's electrical activity. This may be manifest as lower entropy values for EEG signals in schizophrenic patients by comparing with the results for healthy subjects.

5 Conclusions

The primary aim of this study was to investigate the potential application of the symbolic entropy for analyzing the short EEG data collected from both healthy and schizophrenic subjects, and to distinguish the differences of the nonlinearities of the underlying dynamical system. The symbolic entropy proposed in this paper provides a new approach for quantifying the inherent complexity of chaotic sequence of EEG signals. The symbolic dynamics was estimated to characterize the dynamical peculiarity of the EEG data set of two kinds of subjects. Several sequences generated from different chaotic systems were also examined to evaluate and compare their performance of measuring the complexity. The experimental results with both simulated data

and real EEG data demonstrate that the symbolic entropy can effectively distinguish the complexity of difference chaotic series. Our results also indicated that the symbolic entropy is superior to the traditional entropy methods such as BinShan and ApEn. The symbolic entropy may supply us with quantitative characteristics of the EEG signals, which extracts more useful information regarding the nonlinear dynamics of the system than the traditional approaches. Finally, the presented algorithm can be easily performed and computationally simple.

Acknowledgements. This work is supported by the National Natural Science Foundation of china (60571066) and the Natural Science Foundation of Guangdong, respectively.

References

1. Freeman, W.J.: Simulation of Chaotic EEG Patterns with a Dynamic Model of the Olfactory System. *Biological Cybernetics* 56, 139–150 (1987)
2. Freeman, W.J.: Tutorial on Neurobiology: Form Single Neurons to Brain Chaos. *Int. J. Bifurcation and Chaos* 2, 451–482 (1992)
3. Zhang, T., Turner, D.L.: A Visuomotor Reaction Time Task Increases the Irregularity and Complexity of Inspiratory Airflow Pattern in Man. *Neuroscience Lett.* 297, 41–44 (2001)
4. Gallez, D.: Predictability of Human EEG: a Dynamical Approach. *Biological Cybernetics* 64, 381–391 (1991)
5. Roschke, J.: The Dimensionality of Human's Electroencephalogram during Sleep. *Biological Cybernetics* 64(91), 307–313
6. Werner, L.: Dimensional Analysis of the Human EEG and Intelligence. *Neuroscience Letters* 143, 10–14 (1992)
7. Zhang, T., Johns, E.J.: Chaotic Characteristics of Renal Nerve Peak Interval Sequence in Norm Intensive and Hypertensive Rats. *Clin. Exp. Pharmacol Physiol.* 25, 896–903 (1998)
8. Pincus, S.M.: Approximate Entropy as a Measure of System Complexity. *Proc. Natl. Acad. Sci.* 88, 2297–2301 (1991)
9. Pincus, S.M.: Quantification of Evolution from Order to Randomness in Practical Time Series Analysis. *Methods Enzymol* 240, 68–89 (1994)
10. Pincus, S.M.: Approximate Entropy (ApEn) as a Complexity Measure. *Chaos* 5, 110–117 (1995)
11. Joydeep, B.: Complexity Analysis of Spontaneous EEG. *Acta Neurobiol. Exp.* 60, 495–501 (2000)
12. Cysarz, D., et al.: Irregularities and Nonlinearities in Fetal Heart Period Time Series in the Course of Pregnancy. *Herzschr Elektrophys* 11, 179–183 (2000)
13. Cysarz, D., et al.: Entropies of Short Binary Sequences in Heart Period Dynamics. *Am. J. Physiol. Heart Circ. Physiol.* 278, H2163–H2172 (2000)
14. Alligood, K.T., Sauer, T.D., Yorke, J.A.: *Chaos-an Introduction to Dynamical Systems.* Springer, Heidelberg (1996)
15. Francis, C.M.L., Chi, K.T.: Co-Existence of Chaos-Based Communication Systems and Conventional Spread-Spectrum, Communication Systems. In: *International Symposium on Nonlinear Theory and its Applications*, vol. 7, pp. 107–110 (2002)

Phase Self-amending Blind Equalization Algorithm Using Feedforward Neural Network for High-Order QAM Signals in Underwater Acoustic Channels

Yasong Luo, Zhong Liu, Pengfei Peng, and Xuezhi Fu

Electronics Engineering College, Naval University of Engineering,
Wuhan 430033, China

Abstract. Complex-valued and non-constant modulus signals are widely used in modern high-speed underwater acoustic communication systems. Based on this environment, a complex-valued blind equalization algorithm using feedforward neural network is brought forward. Aiming at the defects that traditional constant modulus equalization algorithm can't rectify the phase deflection, the cost function is reformed and also a new modified constant modulus algorithm is given. Besides, the new algorithm is improved by introducing the square decision technique to achieve better convergence speed and less gurgitation. The results of simulation show that this new equalization algorithm not only has the ability of phase self-amending, but also performs better than traditional algorithm in the ability and speed of convergence in high order QAM communication systems.

Keywords: Underwater acoustic communication, Feedforward neural network, Blind equalization, Multipath effect.

1 Introduction

Multipath effect is a big problem that needs to be solved in the underwater acoustic communication field, because it can produce serious inter-symbol interferences which will lead to great descent of the communication quality. In order to get over the bad influences caused by multipath effect, adaptive equalization algorithm depending on training sequences can be used to alleviate the inter-symbol interferences. But as shown in [1], using training sequences also brings the time delay of underwater acoustic communication system and the waste of frequency resource which result in low efficiency in many practical situations.

Blind equalization algorithm does not need the training sequences, so it has better channel utilization ratio and is more suitable for underwater acoustic channels whose frequency resource is limited. From the year 1975, different forms of blind equalization algorithms have been put forward and among these algorithms, the blind equalization algorithm using neural network can not only equalize the minimum phase channel but also equalize the non-minimum phase channel and the non-linear channel as shown in [2]. Just because of the wide applicability, applying the neural network technology to realize channel equalization has been a research emphasis of the underwater acoustic communication field in recent years.

Aiming at the situation that complex-valued and non-constant modulus signals are widely used in high-speed underwater acoustic communication systems, a blind equalization algorithm based on feedforward neural network which is suitable for complex-valued signals is given. In order to make the algorithm have the ability of phase self-amending and get over the problems of low convergence speed and big gurgitation when constant modulus equalization algorithm is used to equalize non-constant modulus signals as shown in [3], a modified constant modulus blind equalization algorithm combined with the square decision technique based on feedforward neural network is elaborated. Results of simulation show that compared with the traditional algorithm, this new algorithm has stronger ability of antimultipath interference and faster convergence speed, besides it can rectify phase's random rotate which makes the whole system more credible.

2 Basic Theory of Complex-valued Blind Equalization Based on Feedforward Neural Network

Equivalent baseband model of the blind equalization based on feedforward neural network is illuminated in Fig.1.

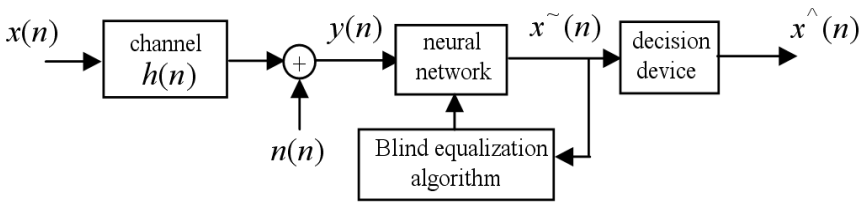


Fig. 1. Equivalent baseband model of the blind equalization based on feedforward neural network

$x(n)$ is the independently uniformly distributed signal sequence sent from source, $h(n)$ is the baseband impulse response of the underwater acoustic channel, and $n(n)$ is the noise sequence. The signal $y(n)$ gotten by the receiver is taken as the input of neural network where $y(n)$ is processed in order to get rid of multipath effects and the output of neural network $\tilde{x}(n)$ is sent into the decision device in order to get $\hat{x}(n)$ which is the estimate of original signal $x(n)$. During the whole process of communication, the weight coefficients of neural network are adjusted continually according to a certain blind equalization algorithm using the input $y(n)$, the output $\tilde{x}(n)$ and the decision value $\hat{x}(n)$. By this way the equalizer can track the changes of underwater acoustic channel and get over the multipath effects which result in bad influences to the underwater acoustic communication system efficiently.

Cybenko has proved in [4] that the feedforward neural network which contains just one hidden layer can approximate any continuous function within any precision. So the

neural network in Fig.1 can be a simple three levels feedforward neural network just as Fig.2 illuminates. $w_{ij}(n)$ is the connective coefficients between the input level and the hidden layer, and $w_j(n)$ are the connective coefficients between the hidden layer and the output level. i represents the nerve cell of the input level whose values can be from 0 to m and j represents the nerve cell of the hidden level whose values can be from 0 to l .

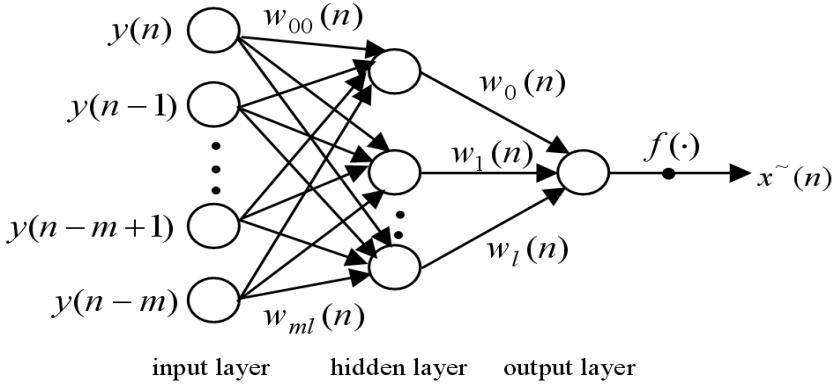


Fig. 2. Structure of three levels feedforward neural network

The signal sent into the input level is a vector $[y(n), y(n-1) \dots y(n-m)]$ whose dimension is $m+1$. $u_j(n)$ is the input of the hidden layer and $s_j(n)$ is its output, and also $v(n)$ and $x^{\sim}(n)$ are the input and output of the output layer respectively.

Because most of the underwater acoustic communication systems use complex-valued signals to transmit information, traditional feedforward neural network algorithm needs to be improved in order to possess the ability to process complex-valued signals. As we know, complex-valued signals are composed of the real part and the imaginary part, so when the complex-valued signals are sent to a nerve cell, its real part and imaginary part can be processed by the non-linear transfer function $f(\cdot)$ respectively and then be combined into a new complex-valued signal as the output of the nerve cell. State equation of the three levels feedforward neural network is represented in the form,

$$\begin{cases} u_j(n) = \sum_i w_{ij}(n) y(n-i) \\ s_j(n) = f[u_{j,R}(n)] + j \cdot f[u_{j,I}(n)] \\ v(n) = \sum_j w_j(n) s_j(n) \\ x^{\sim}(n) = f[v_{j,R}(n)] + j \cdot f[v_{j,I}(n)] \end{cases} \quad (1)$$

where subscripts R and I represent the real part and the imaginary part of a complex number.

3 Complex-valued Blind Equalization Algorithm Based on Feedforward Neural Network That Possesses the Ability of Phase Self-amending

In [5], the cost function adopted by traditional constant modulus blind equalization algorithm based on feedforward neural network(FNN/CMA) is shown in (2), where R_{CM} is a constant lying on precognition features of the original signal $x(n)$ and can be written in the form of formula (3).

$$J_{CM} = [|x^{\sim}(n)|^2 - R_{CM}]^2 / 2 \quad (2)$$

$$R_{CM} = E[|x(n)|^4] / E[|x(n)|^2] \quad (3)$$

From formula (2), it is easy to notice that the constant modulus blind equalization algorithm just uses the amplitude information of signals but not embody the phase information. So random rotates of phase will appear at the output end of the equalizer which causes the phenomenon of phase ambiguity. This phase ambiguity will lead to some misjudgments in the decision device, or even make the whole communication process fail when the extent of ambiguity is serious. In order to solve this problem, J_{CM} , the cost function of constant modulus blind equalization algorithm, is divided into two parts: the real part and the imaginary part. Each part is processed by the constant modulus algorithm respectively and then a new modified constant modulus equalization algorithm (FNN/MCMA) is brought forward whose cost function is represented in the form of (4) where R_R and R_I are constants lying on precognition features of the real part and the imaginary part of original signal $x(n)$. In this way, the cost function J_{MCM} contains not only the amplitude information but also the phase information of signals which makes the new blind equalization algorithm possess stronger ability of phase self-amending.

$$\begin{cases} J_{MCM} = \{ [|x_R^{\sim}(n)|^2 - R_R]^2 + [|x_I^{\sim}(n)|^2 - R_I]^2 \} / 2 \\ R_R = E[|x_R(n)|^4] / E[|x_R(n)|^2] \\ R_I = E[|x_I(n)|^4] / E[|x_I(n)|^2] \end{cases} \quad (4)$$

In the complex-valued neural network system, the coefficients $w(n)$ are also complex. So according to the method of steepest descent, the formula of iteration for $w(n)$ can be represented in the form of formula (5).

$$w(n+1) = w(n) - \mu \cdot \frac{\partial J_{MCM}}{\partial w(n)} = w(n) - \mu \cdot \left[\frac{\partial J_{MCM}}{\partial w_R(n)} + j \cdot \frac{\partial J_{MCM}}{\partial w_I(n)} \right] \quad (5)$$

Combined formula (5) with (1) and (4), the formula of iteration for coefficients $w_j(n)$ in the modified constant modulus blind equalization algorithm based on feedforward neural network can be written in the form of formula (6), where $f'(x)$

represents the derivative of the transfer function $f(x)$, μ is the iterative step, $e_{j,R}$ and $e_{j,I}$ are error signals for adjusting $w_j(n)$, * means getting the conjugate value of a complex number.

$$\begin{cases} w_j(n+1) = w_j(n) - 2\mu s_j^*(n)[e_{j,R} + j \cdot e_{j,I}] \\ e_{j,R} = (|x_R^-(n)|^2 - R_R)f[v_R(n)]f'[v_R(n)] \\ e_{j,I} = (|x_I^-(n)|^2 - R_I)f[v_I(n)]f'[v_I(n)] \end{cases} \quad (6)$$

The formula of iteration for $w_{ij}(n)$, connection weights between the input layer and the hidden layer, can be written in the form of formula (7), where $e_{ij,R}$ and $e_{ij,I}$ are the error signals for adjusting $w_{ij}(n)$.

$$\begin{cases} w_{ij}(n+1) = w_{ij}(n) - 2\mu y^*(n-i)[e_{ij,R} + j \cdot e_{ij,I}] \\ e_{ij,R} = [|x_R^-(n)|^2 - R_R]f[v_R(n)]f'[v_R(n)] \cdot \{w_{j,R}(n)f'[u_{j,R}(n)] - j \cdot w_{j,I}(n)f'[u_{j,I}(n)]\} \\ e_{ij,I} = [|x_I^-(n)|^2 - R_I]f[v_I(n)]f'[v_I(n)] \cdot \{w_{j,R}(n)f'[u_{j,I}(n)] - j \cdot w_{j,I}(n)f'[u_{j,R}(n)]\} \end{cases} \quad (7)$$

The new algorithm (5)~(7) can not only realize blind equalization for the underwater acoustic channels, but also rectify phase deflection automatically for the new cost function J_{MCM} , which gets over the problem of decision misjudgments caused by the phenomenon of phase ambiguity.

From the error signals in formula (6) and (7), it can be noticed that no error signal will approximate 0 at any signal point when the constant modulus equalization algorithm is used to equalize non-constant modulus signals, such as 16QAM and so on. This character makes the whole algorithm have low convergence speed and big gurgitation which can be validated in the fourth part. In order to overcome the problem, the new blind equalization algorithm elaborated above is improved by introducing the square decision technique. The square decision technique divides the signal constellations of 16QAM into two foursquare regions which are equidistant in both synchronized and orthogonal directions, just as Fig.3 illuminates. Before each iteration, $x^-(n)$, the output of equalizer, is firstly checked to make sure the square region i it belongs to and then use formula (8) to replace R_R and R_I in the original algorithm, where S_{Ri} and S_{Qi} are respectively the synchronized component and the orthogonal component of the signal point who has the biggest amplitude in square region i . By this way, the error signals of the blind equalization algorithm do not equal 0 only at a few signal points which quickens convergence speed and reduces gurgitation of the whole algorithm.

$$\begin{cases} R_R = S_{Ri}^4 / S_{Ri}^2 = S_{Ri}^2 \\ R_I = S_{Qi}^4 / S_{Qi}^2 = S_{Qi}^2 \end{cases} \quad i = 1, 2 \quad (8)$$

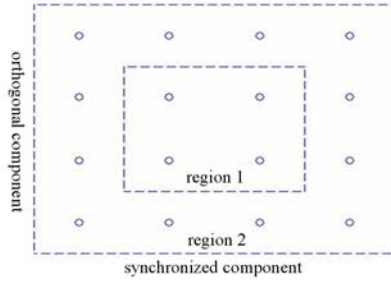


Fig. 3. Constellations of the 16QAM communication system

4 Simulation and Results Analysis

The performance of the modified constant modulus blind equalization algorithm combined with the square decision technique based on complex-valued neural network (FDD/MCMA/SD) is validated by computer simulation and compared with the algorithms FDD/CMA and FDD/MCMA. The mean square error as shown in [6] can be calculated in the form of formula (9),

$$MSE(k) = [h_\delta - CW^*(k)]^H [h_\delta - CW^*(k)] \sigma_s^2 + W^H(k)W(k) \sigma_n^2 \tag{9}$$

where σ_s^2 is the variance of signals sent from source, σ_n^2 is the common noise spectral density when the noises are assumed to be white, C is the underwater acoustic channel correlation matrix, $W(n)$ is the equivalent filter coefficient of the neural network when the time is n , and $h_\delta = [0, \dots, 0, 1, 0, \dots, 0]^T$ is the idea joint impulse response of underwater acoustic channel and equalization filter where the nonzero coefficient is only in the δ th position.

16QAM which is a non-constant modulus signal is used to check the performances of each algorithm and a deep-sea channel model in [7] is cited. The baseband impulse response of underwater acoustic channel h is $[0.3122, -0.104, 0.8908, 0.3134]$, which is a mixed-phase system. The structure of feedforward neural network adopted for equalization is $[7, 3, 1]$ and its coefficients w_{ij} and w_j are initialized by random numbers whose absolute values are less than 0.5. The step size μ equals 0.0001 and the transfer function can be represented in the form of formula (10), which accords with the characteristic demands elaborated in [8].

$$f(x) = x + 0.001 \cdot \sin(\pi x) \tag{10}$$

Fig.4 shows the constellations of 16QAM signals after equalization using FDD/CMA and FDD/MCMA respectively when SNR equals 20dB. It is obvious that constellations of the algorithm FDD/MCMA is clearer than that of FDD/CMA. So the algorithm FDD/MCMA not only equalizes the underwater acoustic channel better than FDD/CMA, but also rectifies the phase deflection that exists when using the algorithm

FDD/CMA, just as Fig.4(b) illuminates. So FDD/MCMA can track the random rotate of phase and keep signals at its original position, which avoids the bad influences caused by decision misjudgments.

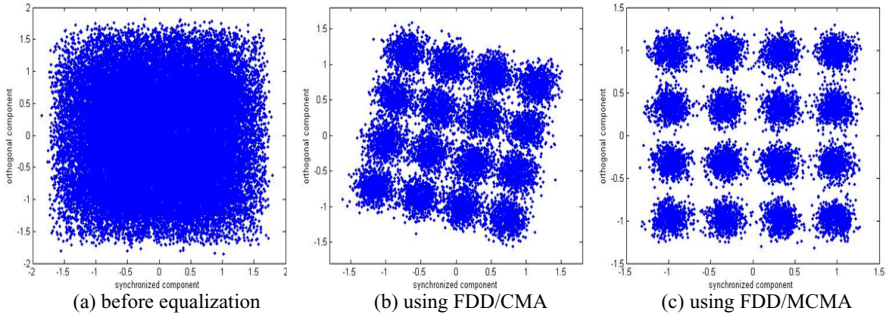


Fig. 4. Constellations of the 16QAM communication system

Fig.5 shows the MSE curves of each algorithm when SNR equals 20dB. It is clear that FDD/MCMA and FDD/MCMA/SD achieve better in the ability of MSE convergence than FDD/CMA which means that modified constant blind equalization algorithm possesses stronger ability of antimultipath interference. Then compared with FDD/MCMA/SD, the algorithm FDD/MCMA has slower convergence speed and bigger gurgitation, that is because its error signals do not equal 0 at any expected signal point which makes the algorithm adjust its coefficients continuously. While FDD/MCMA/SD introduces the square decision technique that makes its error signal equal 0 at most signal points, so it can keep the adjustments more precious and reach the steady state more quickly.

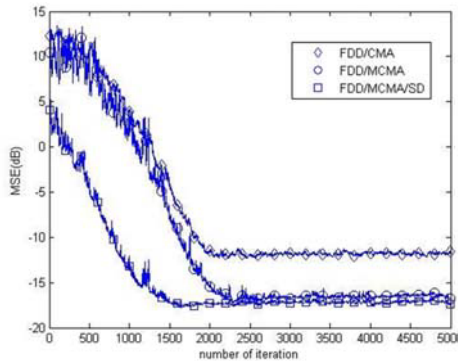


Fig. 5. MSE curves for FDD/CMA, FDD/MCMA and FDD/MCMA/SD (SNR=20dB)

5 Conclusion

Neural network technology has its unique superiority to equalize underwater acoustic channel. Based on the traditional constant modulus blind equalization algorithm, this paper gives a phase self-amending blind equalization algorithm using feedforward neural network for complex high-order 16QAM signals. Aiming at the defects of slow convergence speed and big gurgitation when constant modulus blind equalization algorithm is applied to non-constant modulus signals, this algorithm is improved by introducing the square decision technique which forms a new blind equalization algorithm, namely FDD/MCMA/SD. The results of simulation show that FDD/MCMA/SD not only possesses the ability of phase self-amending, but also performs better than traditional algorithm in the ability and speed of convergence which equalizes underwater acoustic channels more efficiently and keeps the whole communication system steady and credible.

References

1. Kilfoyle, D.B., Baggeroer, A.B.: The State of the Art in Underwater Acoustic Telemetry. *IEEE. J. Oceanic Eng.* 25, 6–25 (2000)
2. Cheolwoo, Y., Daesik, H.: Nonlinear Blind Equalization Schemes Using Complex-valued Multilayer Feed forward Neural Networks. *IEEE Trans. on Signal Processing* 9, 1442–1455 (1998)
3. Zheng, Y.Q., Li, P., Zhang, Z.R.: Dual-mode Blind Equalization Algorithm for Multi-level QAM Modulation Based on Sign-CMA. *Journal of China Institute of Communication* 25, 155–159 (2004)
4. Cybeako, G.: Approximations by Superposition of a Sigmoidal Function. *Math. Control System Signals* 2, 303–314 (1989)
5. Chatha, H.S., Kumar, A., Bahl, R.: Simulation Studies of Underwater Communication System in Shallow Oceanic Channel. In: *Oceans 2002 MTS/IEEE*, vol. 4, pp. 2401–2408 (2002)
6. Johnson, J., Schniter, P., Endres, J.T.: Blind Equalization Using the Constant Modulus Criterion. *A Review Proceedings of the IEEE* 10, 1927–1949 (1998)
7. Gomes, J., Barroso, V.: Blind Decision-feedback Equalization of Underwater Acoustic Channels. In: *Oceans 1998 Conference Proceedings*, vol. 2, pp. 810–819 (1998)
8. Mgeorgiou, G., Koutsougeras, C.: Complex Domain Backpropagation. *IEEE Trans. on Circuits and System* 39, 330–334 (1992)

An Adaptive Channel Handoff Strategy for Opportunistic Spectrum Sharing in Cognitive Global Control Plane Architecture

Zhiming Xu^{1,2}, Yu Wang¹, Jingguo Zhu^{1,2}, and Jian Tang^{1,2}

¹ Department of Space Engineering, Academy of Opto-electronics,
Chinese Academy of Sciences,
Beijing 100190, China

² School of Information Science and Engineering,
Graduate University of Chinese Academy of Sciences,
Beijing 100049, China

Abstract. In the cognitive wireless networks, the channel handoff strategy which controls the communication in a channel to migrate to another channel when primary users hope to access directly affects the quality-of-service of cognitive users and primary users. In this paper, we propose a Cognitive Global Control Plane (CGCP) which embraces the controlling of spectrum sensing, channel handoff and node mobility. With the CGCP architecture, an adaptive channel handoff strategy is given out. We analyze the handoff algorithm with Markov chain theory and some tradeoffs between performance and overhead. The results of performance evaluation show that the channel handoff strategy can adaptively and flexibly perform channel handoff.

Keywords: Cognitive networks, Channel handoff, Strategy, Adaptive, Control.

1 Introduction

For the past decades, with the unprecedented proliferation of wireless mobile devices, it is believed that spectrum is getting acute shortage in the near future. The spectrum regulatory bodies have to start rethinking if their static spectrum strategy is suitable. And they racked their brains to explore how to solve this crisis. In traditionally the regulators of the spectrum granted licenses for spectrum utilization with compulsory and detailed transmission guidelines on the one hand, and on the other hand sliced appropriate guardbands between neighbor frequency bands to guarantee elimination of mutual interference. It is very true that this method makes interference mitigation very easy, but the growth of the number of wireless communications and technologies has caused this frequency allocation strategy difficult to continue. However, extensive measurements reported indicate that the static frequency allocation results in a low utilization (only 6%) of the licensed radio spectrum in most of the time [1].

This kind of technology is called Cognitive Radio (CR), which is proposed by J. Mitola in [2]. CR technology promises tremendous advantages over existing wireless spectrum utilizing methods. A paramount feature of the cognitive radio is that it learns from its surroundings and adapts its configurations to better serve the application. It enables the flexible and efficient use of spectrum bands by adapting physical and link layer working characteristics to the real-time conditions of the environment. Though attractive, this immature and plastic technology has its share of challenges. One of the most basically and important is spectrum sensing which can search for white space of spectrum. There are several methods for white space detecting [3] includes pilot detection, energy detection, cyclostationary detection and wavelet based edge detection. Another is cooperative spectrum sensing and beamforming which are distributed sensing different from the techniques described previously. No matter how to get the white spectrum space, we have to utilize them efficient and effective.

Researching for cognitive radio networks are full of challenges for both wireless and networks communications society. While the development of cognitive radio spectrum sensing [4] and MAC protocol[5], especially white space detection, has received considerable attention, the question of how to use these spectrum and how to handoff between multiple separated bands and keep the normal communication is much less well understood, and there is lacks of research on protocols and strategies to control the handoff in separated frequency band. Furthermore, this question attracted even less attention in cognitive radio networks.

I. Budiarto et al. review all the dynamic spectrum accessing techniques in details [3] and provide some insights on important design considerations and requirements when developing spectrum management schemes to realize real-time spectrum usage. In [6], the authors main works focus on negotiated or opportunistic access strategy to implement real-time secondary spectrum accessing. However, they all do not give out the specific spectrum accessing management architecture in mobile cognitive radio networks. Balamuralidhar P. et al. [7] present a generic architecture for a cognitive node with a context driven approach. The idea is hiding the complex management process from users, but the implicit complexity of itself can not be subtracted. And it does not consider the protocols or algorithm which can be used for cognitive radio networks. X. Jing and Dipankar R. propose a Global Control Plane (GCP) [8,9] architecture for cognitive radio network and utilizing the concepts of "global control plane" and data plane. They evaluate three key components of this architecture while they do not take into account the question of how to process channel handoff when the incumbent users come back. There are some works about QoS based channel handoff in heterogeneous wireless multi-hop networks [10,11,12] but they do not solve this problem in cognitive radio networks with spectrum management.

To improve the aforementioned schemes and enable the CR network more adaptive and flexible, in this paper we propose an adaptive channel handoff strategy for Cognitive Radio networks based on the Cognitive GCP (CGCP) architecture improved from GCP. Based on this architecture, we design a channel handoff algorithm integrated with spectrum sensing. The handoff also can

be performed efficiently in mobile scenario as it can adapt to the changing of networks and channels.

The rest of this paper is organized as following. First of all in section 2, we present the CGCP. Following in section 3, proposed adaptive handoff strategy and the algorithm process for channel handoff. Section 4 analyze how the algorithm working and some tradeoffs in generic network scenarios. Section 5 gives the simulation results of evaluating our algorithm. The paper concluded with section 6.

2 Model and Assumption

2.1 The Spectrum Sensing Cognitive Global Control Plane Architecture

The architecture we used here is come from the GCP architecture and is shown in Fig. 1. Enlightened by[8], we also divide the network into separated cognitive and data plane logically. The main difference of our CGCP with GCP is in control plane which also have three important components: *initialization*, *routing* management and *spectrum management*. Among them, the responsibility of initialization is to identify the value of all the prearranged protocol parameters and thresholds at the beginning of appeared in networks. Of course, it also collecting some basic physical and data link layer parameters which are transparent to our handoff because of the existing of Control MAC and Control PHY. The routing management mainly focuses on how to change the route information and which path changed during its working. It can get a path to a specific target through *route request*, and also can know which path changed by periodically *route change reporting*. However, the globe route information is gotten through one-hop *route request broadcast*.

We mostly consider the Spectrum Management (SM) which not only sense idle spectrum but also control utilizing the spectrum (or channels) and channel handoff in the CGCP. Actually, spectrum hole is sensed by PHY/MAC with the controlling of Spectrum Management, which will decide when and how to sense with different methods, parameters and thresholds. During the procedure of handoff, Spectrum Management plays an important role in interacting with

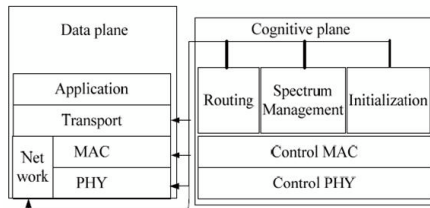


Fig. 1. Cognitive GCP architecture for CR networks. This shows the architecture consisting of *data plane* and *cognitive plane*.

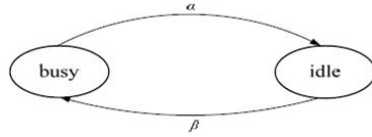


Fig. 2. Channel state transit diagram. It is assumed that at the beginning of each sensing interval, the probability of a channel ch_i changes from busy to idle is α and the probability from idle to busy is β .

control MAC, control PHY and the layers in data plane. In fact, the execution of routing, channel accessing and channel detecting is finished by the related layers in data plane. The network and MAC, PHY has the direct interaction which we call the lower cross-layer interaction. As to the detail of data plane, we will not go further here and you can get more in [8,9] if any interest.

2.2 System Model

We consider a network of N_{CR} Cognitive Users (CUs) opportunistically sharing K (K is a plus integer and $K > 0$) channels, we called data channels, which has been licensed to Primary Users (PUs). Each CU $cu_T, 1 \leq T, R \leq N_{CR}$, can access channel ch_i when all the PUs are not using that channel and can communicate with its target CU cu_R , here $1 \leq i \leq K$. We assume that cu_T is transmitter and cu_R is the receiver here. CUs sense the channel which they want to access in a prearranged periodical interval, namely Δt . A channel is busy when there is a PU using it and is idle if no PU using it. We think that each of the channel changes their state (busy or idle) obeying independent identification distribution (*i.i.d*). Hence, we can model the channel state as a two states first-order Markov chain, as shown in Fig. 2. We suppose that we have a separated channel which is used as Control channel (CH) specially. Naturally, CH will not interfere PU and it will be accessed through competing by CUs.

The periodical interval, t also can be called frame, for channel sensing is defined as a Time Division Multiplex Access (TDMA) channel frame. As in Fig.3, the length of frame t is fixed once you set, but it can be adaptively adjusted with different network situations (In order to adopt different network

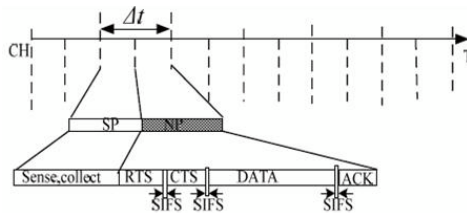


Fig. 3. The components of channel frame. The period Δt can be divided into two parts which are Sensing Phase (SP) and Negotiating Phase (NP).

MAC protocols). In sensing phase, the receiver scans the spectrum to find idle channels that can be utilized. The control channel CH starts to negotiate and setting up data link through IEEE802.11 DCF on CH in the negotiation phase. Once the link is established, CUs use the data channels to transmit application data. As a matter of fact, it only scans spectrum for discovering idle channels by itself during initializing sense phase. After that, CUs cooperatively sense through exchanging their channel state periodically.

Assume that We use energy detection here when we check if PU comes back on the specific channel. During the detection, all the channel transmitting or not is judged by means of received energy on that. It is thought PU comes back if CU finds the received energy is higher than a threshold value (the maximum power during normal communication), and then CU will send this alarm to itself and its neighbors. Obviously, handoff (in section 3) needs to be performed if CU received alarm or interfering and we call the threshold value $thre_{handoff}$. Another, let P_{false} be the probability of sending alarm but there is no PU comes back and P_{miss} be the probability of missing send alarm when PU comes back and leads to collision. Hereafter, we think the neighbor is one hop CU if there is no special emphasis.

3 Adaptive Channel Handoff Strategy

We assume that CU can use N_{using} channels transmitting data simultaneously and need to sense N_{idle} channels put into its idle channel queue in order to assure successful handoff with the probability P_{succ} . We define the P_{succ} as: $P_{succ} = \frac{n_q}{n_{hf}}$, here n_q is the number of channels need to queue in a buffer q for handoff and $n_q = \lceil (1 - \beta)N_{idle} \rceil$, n_{hf} is the total handoff probability for all using channels and $n_{hf} = \lceil \beta N_{using} \rceil$. So we can get the relation between N_{using} and N_{idle}

$$N_{idle} = \lceil \frac{\beta}{1 - \beta} P_{succ} \times N_{using} \rceil. \quad (1)$$

Our solutions to adaptive channel handoff in mobile wireless networks and with a control channel CH includes three phrases: initializing, cooperative sensing and information exchanging, handoff. It is shown in Fig. 4. The first phase is to initialize some parameters and the original states of network and channels. Here the Δt is the handoff delay which is the time of SP+NP in Fig. 3. Get the communication links through the negotiation of CH with other network nodes. After that, the changing of neighbor nodes and used channels need to be update in periodic. So the information is collected through the respective reports in sensing phase. Another task during this phase is keeping the balance of equation (1) so that we can get expected successful handoff when primary users come back. As to spectrum sensing, we have two kinds of sensing which are idle channel sensing and interfering sensing. Channel is idle if we found that the energy is not thermal noise and white noise in the channel. We can know a using channel which is interfered when we found the bit error rate (BER) over the degree which can be put up with for a normal communicating. How to sense

channel state is not described because the main target is how to control but to sense.

Algorithm1. Adaptive channel handoff algorithm

Initializing

01: Initialize $n_q, n_{hf}, N_{using}, n, N = 0$ and $\Delta t = 0$

// Δt is a counter which will be zero in every initialization, increase with the time spent automatically.

02: Network broadcasts neighbor discover requests on CH

03: Sensing and choose N_{using} idle channels for communication, n_q channels for q

// 03 is finished by the cooperation of PHY and SM.

04: If $n_q \leq n_{hf}$

05: Return 02

Periodic sensing phase

06: If PHY received any routing and channel state reports

07: Network updates route table and SM updates channel state in q

08: If $n_q \leq n_{hf}$ // if the idle channel is not enough.

09: Control PHY and MAC choose idle channel

10: If PHY gets any idle channel

11: SM puts it into q

12: Else turn to 08

13: Else turn 14

Handoff phase // the ch_i may be a channel set.

14: If received alarm from PHY on ch_i

15: $n_{hf} = \|ch_i\|, N = N + 1$ // $\|ch_i\|$ is the channel number in ch_i .

16: If $\|q\| = 0$ and all N_{using} channels are interfered

17: $n = n + 1, n_{hf} = \lceil n_{hf} \times GF(n) \rceil$ and turn to 01

18: Else if $\|q\| = 0$ and $n_{hf} > 0$

19: $n = n + 1, n_{hf} = \lceil n_{hf} \times GF(n) \rceil$ and turn to 02

20: Else if $n_{hf} = 0$

21: Turn to 30

For each interfered channel in ch_i

22: Notify the TCP handoff avoiding Congestion Mechanism

23: Choose ch_j from $q, n_q = n_q - 1$

24: If CH is successful in negotiating, set link instead of ch_i

25: If Network finds route for current CU

26: $n_{hf} = n_{hf} - 1$

27: start to transmit data on ch_j and close ch_i

28: Network pushes route and channel to PHY and reports it to neighbors

29: Turn to 14

30: If $\Delta t \geq t_{max}$ // t_{max} is the sensing and Negotiating phase.

31: Turn to 06

End.

When there is any alarm which means some channels are interfered and PUs come back, the algorithm will check q for any alternative channel. We can use a specific metric to sort and choose the channels in q . Next the CH will start to set up new link for current node if the alternative channel is chosen successfully. It is also possible that failing to negotiate with neighbor or get a right route for the current nodes. Or else it will try to do that again until q is empty. We also define a Gain Function (GF)

$$GF(n) = \delta \frac{n}{1 - \beta} \times \frac{n}{N} \quad (2)$$

Here, δ is a accommodating factor and usually equals to 1 (in order to increase or decrease the channel in q , we can adjust it accordingly), n is the times of fail to handoff, N is the total handoff times. In the right of (2), the ratio of n and N denotes the frequency of failing to perform handoff. This function is a greedy punishing function which can adjust n_{hf} so that changes the number of idle channel need to sense.

4 Analysis and Discussion

Cross-layer interaction. There are several aspects of cross-layer interaction in our handoff strategy. Firstly, the spectrum sensing in PHY and spectrum management based on MAC interacting with each other for two goals which embrace monitor the using channel and seize enough idle channels that can be used for handoff. This mechanism can shorten the delay between two layers and the handover spent in communication. Secondly, the interaction of PHY, MAC and network layer can guarantee that any changes on link or route can be updated as soon as possible and report any local route changes to neighbors. So the stale route and link in all nodes can be gotten rid of in time and efficiently. Finally, transport protocols interact with MAC and PHY for controlling the data stream in high-level according to the parameters changes in PHY and MAC. Hence, the handoff can bridge the performance gaps between handoff and normal communication.

Supports for mobility and QoS. In the mobile network scenario, the route and available channel set changes with the moving of nodes. Our algorithm can tackle with this change through periodic route and channel information exchanging. At least they can sense the available channel in its coverage if miss the channel reports. Although Qos for handoff can be provided with different delay time and successful probability, our CGCP can add new QoS managements paralleling with the spectrum management very easily. For example, the supports to real-time data and multimedia stream in heterogeneous networks.

Delay of handoff and performance degrade. Once a channel which is transmitting data is interfered, system has to move to an alternative channel as soon as possible. During this process which is the period from interfered state to successful handoff and we call it handoff delay, data rate will decrease due to

interference. Naturally, the shorter time spend on handoff, the better performance can be obtained. So we must maximize the P_{succ} , and it is an optimizing problem in (3)

$$\operatorname{argmax} \sum_{0 \leq i \leq k} P_{succ}^i = \operatorname{argmax} \frac{1 - \beta}{\beta} \cdot \frac{N_{idle}}{N_{using}} \quad (3)$$

Here the p_{succ}^i is the probability of ch_i successful handoff.

We also consider punishing the idle channel sensing when it failing to perform handoff by using the GF(n) in (2). In other words, we can compute n_q , the number of idle channels in q , generally satisfying the requirement of (1). Once we can not find idle channel for handoff that is meaning handoff failing, then we have to adjust the n_{hf} , using GF(n) multiply the original n_{hf} , before we compute n_q satisfy (1).

Available channel set and handoff frequency. From (3), we also can know that the more channels in idle state, the bigger of p_{succ} . So that more idle channels can be sensed if system has a bigger available channel set and they are *i.i.d.* There is another special situation that is PHY alarming the handoff while there is no idle channel though network has a lot of candidate channels. In this situation, the handoff frequency may be very high because none of them are successful. This happens when the networks are very busy especially the primary users. Once it happens, the cognitive users should stop their communicating with any nodes.

5 Simulation Results

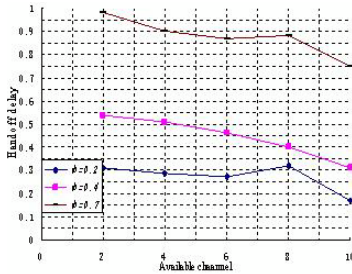
5.1 Simulation

We consider a network with 5 primary users and the number of cognitive users ranges from 2 to 10 in the 500m x 500m simulation area using NS2 and Matlab. Only a transceiver is used in a CU which has a control channel, several data channels which are licensed to primary users and with different channel utilization. We just simulate the PUs' transmitting by a busy signal and without real data. Accessing control channel need not negotiation but control channel has to negotiate the using of data channel with the IEEE802.11 DCF. The available channel set has 10 channels can be disable or enable for comparison. In order to simply, we think 2 nodes are neighbors when there are no others or they can communicate directly. We choose a CU in random moving at the speed of 5m per second during the simulation. The tracks of moving are chosen in random too.

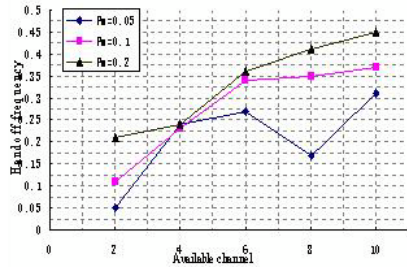
It is evaluated in terms of handoff delay, handoff frequency and successful handoff rate in different available channel set. For comparisons, we let them run in different channel utilizations and all the numerical results in 5.2 are the processed data.

5.2 Results

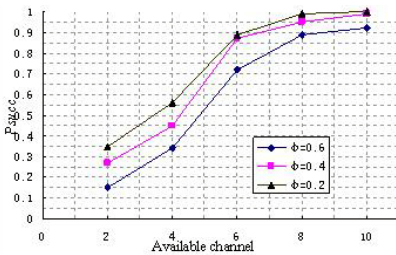
The relations of the number of available channel and handoff delay which is the time spent on handoff are given in Fig. 4(a), and we compare that in different channel utilizations. The y-axis indicates handoff delay which is the ratio of handoff time spent and t_{max} . Obviously, the handoff delay is increasing with the channel utilization (φ) get high while changes very little in different available channels. We compared our algorithm with LSA [8] in our simulating configuration about successful accessing rate (SAR), and the results are in Fig. 4(d). Obviously, our algorithm performs better than LSA, which just considering the link state changes without specific handoff handling measure.



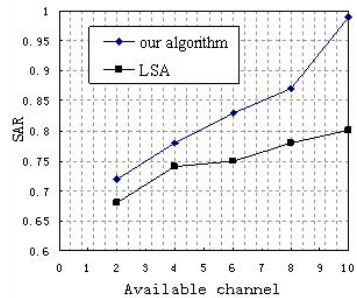
(a) handoff delay with different channels



(b) handoff frequency of mobile nodes



(c) P_{succ} with different available channel



(d) comparison of SAR ($P_m=0.2$)

Fig. 4. Results on the simulation: (a) handoff delay with different channels. (b) handoff frequency of mobile nodes. (c) P_{succ} with different available channel. (d) comparison of successful accessing rate (SAR) ($P_m=0.2$).

As is shown in Fig. 4(b), handoff frequency gets lower with the increasing of available channel. Handoff frequency is computed using the handoff number divides the total periodic sensing number. The handoff frequency fluctuates obviously under different mobility a probability (P_m), due to not only channel gets invalid but also the route gets invalid. It is also shown that the handoff frequency gets high with the increasing of available channel that is because there

is no channels can perform handoff when available channel set is very small. So handoff frequency gets same when the available channel is 4, but the handoff frequency get lower when $P_m = 0.5$ and available channel is 8 because there are enough channels can be used.

In Fig. 4(c), the probability of successful handoff in different available channel number is depicted. Because the more channel the more opportunity we can use in the same channel utilization. So the bigger successful handoff probability can be gained. However, the successful handoff probability is decreasing when channel utilization increasing because the available channel opportunity gets lower.

6 Conclusions

We proposed a CGCP which integrates with spectrum sensing, routing management and channel handoff for cognitive wireless networks. Based on the CGCP architecture, we give out an adaptive channel handoff strategy in detail. In order to maximize successful handoff rate which means smooth the QoS difference during channel handoff, we sense specific idle channels in queue which will substitute the interfered channel. The number of idle channels in queue is determined with the expected successful handoff rate. Finally, the simulations results confirm the effectiveness of our proposed strategy in efficient channel handoff. Though it performs well enough in our simulation, the handoff only can be implemented in deteriorated channel environment with graceful performance degradation. Hence in the future, we will seek better handoff or roaming management scheme, and extend the handoff algorithm in multi-transceivers cognitive radio networks based our CGCP architecture, analyze and verify the performance.

References

1. Mchenry, M.: Spectrum white space measurements. New America Foundation Broadband Forum (2003)
2. Joseph III, M.: Software Radio Architecture. John Wiley & Sons Inc, Chichester (2000)
3. Budiarjo, I., Lakshmanan, M.K., Nikookar, H.: Cognitive Radio Dynamic Access Techniques. *Wireless Pers. Commun.* 45, 293–324 (2008)
4. Dehni, L., Krief, F., Bennani, Y.: Power Control and Clustering in Wireless Sensor Networks. In: *Proceedings of IFIP Med-Hoc-Net 2005. Mediterranean Ad Hoc Networking Conference*, pp. 1–11 (2005) (invited paper)
5. Mehta, V., Zarki, M.E.: An Ultra Wide Band (UWB) Based Sensor Network for Civil Infrastructure Health Monitoring, <http://vip.ics.uci.edu/publications/2004/EWSN2004.pdf>
6. Mcknight, L.W., Howison, J., Bradner, S.: Wireless Grids Distributed Resource Sharing by MobileNomadic and Fixed Devices. *IEEE Internet Computing*, 24–31 (2004)

7. Balamuralidhar, P., Prasad, R.: A Context Driven Architecture for Cognitive Radio Nodes. *Wireless Pers. Commun.* 45, 423–434 (2008)
8. Jing, X., Raychaudhuri, D.: Global Control Plane Architecture for Cognitive Radio Networks. In: *ICC 2007 proceedings*, pp. 6466–6470. ICC (2007)
9. Raychaudhuri, D., Mandayam, N.B., Evans, J.B., et al.: CogNet - An Architectural Foundation for Experimental Cognitive Radio Networks within the Future Internet. In: *Proceedings of MobiArch 2006, PMA* (2006)
10. Cavalcanti, D., et al.: Connectivity opportunity selection in heterogeneous wireless multi-hop networks. *Pervasive and Mobile Computing* 4, 390–420 (2008)
11. Qingyang, S., Jamalipour, A.: A quality of service negotiation-based vertical hand-off decision scheme in heterogeneous wireless systems. *European Journal of Operational Research* 191, 1059–1074 (2008)
12. Gazis, V., Alonistioti, N., Merakos, L.: Toward a generic always best connected capability in integrated WLAN/UMTS cellular mobile networks (and Beyond). *IEEE Wireless Communications* (2005)

A Generalization of the Bent-Function Sequence Construction

Yongbo Xia^{1,2}, Yan Sui³, and Junhao Hu²

¹ The Faculty of Mathematics and Computer Science,
Hubei University, Xueyuan Road 11,
Wuhan 430062, P.R. China

² School of Computer Science,
South-Central University for Nationalities, Minyuan Road 708,
Wuhan 430074, P.R. China

³ Department of Fundamental Courses,
Wuhan Electric Power Technical College, Luoyu Road 189,
Wuhan 430079, P.R. China
xiayongbom@hotmail.com

Abstract. The construction method of bent-function sequences is generalized to construct sequences with low correlation from functions with low maximal Walsh transform. In detail, for any nonlinear function from $F_{2^e}^k$ to F_2 with maximum magnitude spectra A , a family of sequences with period $2^{2ek} - 1$ can be constructed. There are 2^{ek} sequences within a family and the maximum nontrivial auto and cross-correlation values equals $A^2 + 1$. The linear span of these proposed sequences is discussed and the lower bound can be greater than $\binom{ek}{d} 2^d + 2ek$, where $2 \leq d < ek$.

Keywords: bent-function sequences, correlation, nonlinear function, Walsh transform.

1 Introduction

Families of sequences with low cross correlation have a wide range of applications in CDMA communications and cryptography. Using bent function over F_2^m , J.D.Olsen, R.A.Scholtz and L.R.Welch [1] introduced bent-function sequence with optimal correlation and balance property. Later, No [2] generalized the construction method presented in [1] and obtained a family of generalized binary bent sequences with optimal correlation and balance property. However, up to now only a few kinds of bent functions are known and there exists a lot of nonlinear functions with low Walsh transform. In this paper, the construction method of bent-function sequences is generalized to construct sequences with low correlation from functions with low maximal Walsh transform. The new proposed families of sequences also have low correlation and balance property. The linear spans of these proposed sequences are also discussed.

The remaining part of this paper is organized as follows. Section 2 introduces preliminaries that are used in this paper. Section 3 gives the sequence construction method, which includes the bent sequences construction as a special case and section 4 discusses their linear span. Concluding remarks are in Section 5.

2 Preliminaries

Let F_{2^m} be the finite field with 2^m elements, and we always assume that $m = ek$ for some positive integers e and k . The trace function $tr_e^m(\cdot)$ from F_{2^m} to its subfield F_{2^e} is defined by

$$tr_e^m(x) = \sum_{i=0}^{k-1} x^{2^{ei}}, \quad x \in F_{2^m}.$$

The properties of trace function can be found in [3].

Let $F_{2^e}^k$ be the k -dimensional vector space over F_{2^e} and $f(x)$ be a function from $F_{2^e}^k$ to F_2 . The Walsh transform is defined as follows

$$\widehat{f}(\lambda) = \sum_{x \in F_{2^e}^k} (-1)^{f(x) + tr_1^e(\lambda \cdot x)}, \tag{1}$$

where $\lambda \cdot x = \lambda_1x_1 + \lambda_2x_2 + \dots + \lambda_kx_k$ means the inner product of λ and x in $F_{2^e}^k$. Let $g(x)$ be also a function from $F_{2^e}^k$ to F_2 , the following properties of Walsh transform can be easily derived.

(i) Inversion:

$$(-1)^{f(x)} = \frac{1}{2^m} \sum_{\lambda \in F_{2^e}^k} \widehat{f}(\lambda) (-1)^{tr_1^e(\lambda \cdot x)}. \tag{2}$$

(ii) Parseval's Relation:

$$\sum_{x \in F_{2^e}^k} (-1)^{f(x) + g(x)} = 2^{-m} \sum_{\lambda \in F_{2^e}^k} \widehat{f}(\lambda) \widehat{g}(\lambda). \tag{3}$$

For every function $f(x)$ from $F_{2^e}^k$ to F_2 , $\max_{\lambda \in F_{2^e}^k} |\widehat{f}(\lambda)|$ is called its maximum magnitude spectra and the value

$$N_f = 2^{m-1} - 2^{-1} \cdot \max_{\lambda \in F_{2^e}^k} |\widehat{f}(\lambda)|$$

is used to measure the nonlinearity of f . Because of Parseval's relation (3), N_f is upper bounded by $2^{m-1} - 2^{m/2-1}$. This bound is tight for the every even m and the following function can achieve it.

Definition 1. [4,5] Let m be even. The function $f(x) : F_{2^e}^k \rightarrow F_2$ is bent if $|\widehat{f}(\lambda)| = 2^{\frac{m}{2}}$ for all $\lambda \in F_{2^e}^k$.

Definition 2. [6,7] Let m be odd. The function $f(x) : F_{2^e}^k \rightarrow F_2$ is r -plateaued function if $|\widehat{f}(\lambda)| \in \{0, 2^{\frac{2m-r}{2}}\}$ for all $\lambda \in F_{2^e}^k$.

The plateaued functions includes partially-Bent functions; see [6,7,8,9,10].

Let \mathcal{S} be a family of M binary sequences of period N

$$\mathcal{S} = \{ \{s_i(t)\}_{t=0}^{N-1} \mid 0 \leq i < M \}.$$

Then, the correlation function between $\{s_i(t)\}_{t=0}^{N-1}$ and $\{s_j(t)\}_{t=0}^{N-1}$ is

$$R_{i,j}(\tau) = \sum_{t=0}^{N-1} (-1)^{s_i(t)+s_j(t+\tau)}, \quad 0 \leq i, j < M, \quad 0 \leq \tau < N. \tag{4}$$

The maximum correlation of \mathcal{S} is defined by

$$C_{max} = \max_{i \neq j \text{ or } \tau \neq 0} \{ |R_{i,j}(\tau)| \}.$$

If there exist a constant c such $C_{max} \leq c\sqrt{N}$, we say that \mathcal{S} has low cross correlation.

3 A Generalized Method of Sequence Construction

In this section, we will give a basic method of sequence construction and analyze its properties on cross correlation.

From now on, we will use the following notations frequently:

- $n = 2m = 2ek$;
- α : a primitive element of F_{2^n} ;
- $\sigma \in F_{2^n} \setminus F_{2^m}$: a fixed element in F_{2^n} but not in F_{2^m} ;
- $\{\beta_1, \beta_2, \dots, \beta_k\}$: a basis of F_{2^m} over F_2 ;
- $\theta, \mu \in F_{2^m}^*$: two nonzero elements in F_{2^m} .

Let $f(x)$ be a function from $F_{2^e}^k$ to F_2 with the maximum magnitude spectra equal to $\max_{\lambda \in F_{2^e}^k} \{ |\widehat{f}(\lambda)| \} = A$. Define a family of binary sequences as in the follows

$$\begin{aligned} \mathcal{S}_\theta &= \{ \{s_{\eta, \theta}(t)\}_{t=0}^{2^n-2} \mid \eta \in F_{2^m} \}, \\ s_{\eta, \theta}(t) &= f(L(\alpha^t)) + tr_1^n((\eta + \theta\sigma)\alpha^t), \end{aligned} \tag{5}$$

where $L(x) = (tr_e^n(\beta_1x), \dots, tr_e^n(\beta_kx))$ is the linear mapping from F_{2^n} to F_2^k .

The following two propositions are critical for determining the correlation of sequences in \mathcal{S}_θ .

Proposition 1. Let σ be a fixed element in $F_{2^n} \setminus F_{2^m}$ and $H_\mu = \{ \mu\sigma + \delta \mid \delta \in F_{2^m} \}$. Then, for any given $\theta, \mu \in F_{2^m}^*$ and any $\tau \in \{0 \leq \tau < 2^n - 1 \mid \alpha^\tau \neq \mu\theta^{-1}\}$,

$$|H_\theta \cap \alpha^\tau H_\mu| = \begin{cases} 0, & \text{if and only if } \alpha^\tau \in F_{2^m}, \\ 1, & \text{if and only if } \alpha^\tau \notin F_{2^m}. \end{cases}$$

Proof. First, Since $\{1, \sigma\}$ is a basis of F_{2^n} over F_{2^m} , then $\varphi(x_1, x_2) = \frac{\sigma+x_1}{\sigma+x_2}$ is a one-to-one mapping from $G = \{(x_1, x_2) \mid (x_1, x_2) \in F_{2^m}^2, x_1 \neq x_2\}$ to $F_{2^n} \setminus F_{2^m}$.

Then, it can be verified that $H_\theta = \theta\mu^{-1}H_\mu$ and thus

$$H_\mu \cap \alpha^\tau H_\theta = H_\mu \cap \theta\mu^{-1}\alpha^\tau H_\mu.$$

Let $\lambda \in H_\mu \cap \alpha^\tau H_\theta = H_\mu \cap \theta\mu^{-1}\alpha^\tau H_\mu$. Then there exist $\beta_1, \beta_2 \in F_{2^m}$ such that

$$\begin{cases} \lambda = \mu\sigma + \beta_1, \\ \lambda\alpha^{-\tau}\mu\theta^{-1} = \mu\sigma + \beta_2. \end{cases}$$

Therefore,

$$\begin{cases} \lambda = \theta\sigma + \beta_1, \\ \alpha^\tau\theta\mu^{-1} = \frac{\theta\mu + \beta_1}{\theta\mu + \beta_2}. \end{cases}$$

Since $\tau \in \{0 \leq \tau < 2^n - 1 \mid \alpha^\tau \neq \mu\theta^{-1}\}$, by the fact $\varphi(x_1, x_2)$ is a bijective from G to $F_{2^n} \setminus F_{2^m}$, we know (β_1, β_1) is uniquely determined by τ . Thus, λ is uniquely determined by τ . If $\alpha^\tau \in F_{2^m}$, there is no (β_1, β_1) for the above equations holding. Accordingly, $|H_\theta \cap \alpha^\tau H_\mu| = 0$. If $\alpha^\tau \notin F_{2^m}$, there is exactly one (β_1, β_1) for the above equations holding. Accordingly, $|H_\theta \cap \alpha^\tau H_\mu| = 1$. When τ range over $\{0 \leq \tau < 2^n - 1 \mid \alpha^\tau \neq \mu\theta^{-1}\}$, there are $2^m - 2$ τ such that $\alpha^\tau \in F_{2^m}$ and $2^n - 2^m - \tau$ such that $\alpha^\tau \notin F_{2^m}$. This completes the proof.

Proposition 2. Let $F_\eta(x) = f(L(x)) + tr_1^n((\eta + \theta\sigma)x)$ with $\eta \in F_{2^m}$ and $H_\theta = \{\theta\sigma + \delta \mid \delta \in F_{2^m}\}$, then

$$\widehat{F}_\eta(\lambda) = \begin{cases} 2^m \widehat{f}(\gamma), & \text{if } \lambda \in H_\theta, \\ 0, & \text{if } \lambda \notin H_\theta, \end{cases}$$

for some $\gamma = (\gamma_1, \dots, \gamma_k) \in F_{2^e}^k$ which satisfies $\sum_{i=1}^k \gamma_i \beta_i + \eta + \theta\sigma = \lambda$.

Proof

$$\begin{aligned} \widehat{F}_\eta(\lambda) &= \sum_{x \in F_{2^n}} (-1)^{f(L(x)) + tr_1^n((\eta + \theta\sigma + \lambda)x)} \\ &= \sum_{x \in F_{2^n}} \frac{1}{2^m} \sum_{\gamma \in F_{2^e}^k} \widehat{f}(\gamma) (-1)^{tr_1^n(\sum_{i=1}^k \gamma_i \beta_i x)} (-1)^{tr_1^n((\eta + \theta\sigma + \lambda)x)} \\ &= \frac{1}{2^m} \sum_{\gamma \in F_{2^e}^k} \widehat{f}(\gamma) \sum_{x \in F_{2^n}} (-1)^{tr_1^n((\eta + \theta\sigma + \lambda + \sum_{i=1}^k \gamma_i \beta_i)x)} \end{aligned}$$

with $\gamma = (\gamma_1, \dots, \gamma_k) \in F_{2^e}^k$. Note that $\{\theta\sigma + \eta + \sum_{i=1}^k \gamma_i \beta_i \mid \gamma \in F_{2^e}^k\} = H_\theta$ and

$\sum_{x \in F_{2^n}} (-1)^{tr_1^n((\eta + \theta\sigma + \lambda + \sum_{i=1}^k \gamma_i \beta_i)x)} = 0$ if and only if $\lambda \notin H_\theta$. Thus, the proof is finished.

Theorem 1. The sequence set \mathcal{S}_θ constructed in Eq. (5) is a family of 2^m binary sequences of period $2^{2^m} - 1$ with maximum correlation equaling $A^2 + 1$ and balance property.

Proof. Let $s_{\eta,\theta}(t), s_{\eta',\theta}(t) \in \mathcal{S}_\theta$. First, we prove the balance property of the sequences in the family \mathcal{S}_θ . The imbalance of the sequence $s_{\eta,\theta}(t)$, denoted by $I(s_{\eta,\theta}(t))$, is calculated as follows

$$\begin{aligned} I(s_{\eta,\theta}(t)) &= \sum_{t=0}^{2^n-2} (-1)^{f(L(\alpha^t))+tr_1^n((\eta+\sigma)\alpha^t)} \\ &= \widehat{F}_\eta(0) - 1, \end{aligned}$$

where $F_\eta(x) = f(L(x) + tr_1^n((\eta + \theta\sigma)x))$. Since $0 \notin H_\theta$, by proposition 2, we have $\widehat{F}_\eta(0) = 0$. Thus, $I(s_{\eta,\theta}(t)) = -1$. Hence, every sequence in in the family \mathcal{S}_θ is balanced.

We consider the cross correlation in the following two cases.

(i) If $1 \leq \tau \leq 2^n - 2$, then by the properties (2) and (3) of Walsh transformation, the cross correlation between the sequences $s_{\eta,\theta}(t)$ and $s_{\eta',\theta}(t)$ can be written as

$$\begin{aligned} R_{\eta,\eta'}(\tau) &= -1 + \sum_{x \in F_{2^n}} (-1)^{F_\eta(x)+F_{\eta'}(\alpha^\tau x)} \\ &= -1 + 2^{-n} \sum_{\lambda \in F_{2^n}} \widehat{F}_\eta(\lambda)\widehat{F}_{\eta'}(\alpha^{-\tau}\lambda) \\ &= -1 + \sum_{\lambda \in H_\theta \cap \alpha^\tau H_\theta} \widehat{f}(\beta_1)\widehat{f}(\beta_2) \\ &\leq 1 + |H_\theta \cap \alpha^\tau H_\theta|A^2 \\ &\leq 1 + A^2. \end{aligned}$$

(ii) If $\tau = 0$ and $\eta \neq \eta'$, it is easily derived that $R_{\eta,\eta'}(0) = -1$. The proof is finished.

Remarks. (i) Theorem 1 generalizes the construction method of bent-function sequences. For any given nonlinear function $f(x)$, Theorem 1 shows that the maximal correlation of \mathcal{S}_θ is only dependent on the maximum magnitude spectra of $f(x)$.

(ii) If $n = 0 \pmod{4}$, there exist bent functions $f(x)$ from $F_{2^e}^k$ to F_2 , i.e., $A = 2^{m/2}$. Then, the sequences constructed in Eq. (5) are the Bent-function sequences presented in [6,7,10], which have optimal correlation property and large linear span.

(iii) If $n = 2 \pmod{4}$, there exist many nonlinear functions from $F_{2^e}^k$ to F_2 with maximum magnitude spectra $A \leq 2^{(m+1)/2}$. Thus, using the method providing in Theorem 1, we can also construct families of sequences with low correlation.

For $n = 2 \pmod{4}$, the following examples present some nonlinear functions from $F_{2^e}^k$ to F_2 with maximum magnitude spectra equal to $A = 2^{(m+1)/2}$.

Example 1. Let $m = ek$ be odd and $k = 2t + 1$. Suppose $p(x)$ is a $(e - 1)$ -plateaued function over F_{2^e} [11] and $\pi_i(x_1, \dots, x_t)$ is a permutation of x_1, \dots, x_t for $i = 1, \dots, t$. Let $f(x)$ be a function from $F_{2^e}^k$ to F_2 given by

$$\begin{aligned} f(x) &= tr_1^e(\pi_1(x_1, \dots, x_t)x_{t+1} + \pi_2(x_1, \dots, x_t)x_{t+2} + \\ &\quad \dots + \pi_t(x_1, \dots, x_t)x_{2t} + g(x_1, \dots, x_t)) + p(x_k). \end{aligned}$$

Then, by a direct calculation, $f(x)$ is a $(m - 1)$ -plateaued function over F_2^k , i.e., the maximum magnitude spectra of $f(x)$ is $2^{(m+1)/2}$.

Example 2. [12] Assume that m is odd. Let $f_0(\mathbf{x})$ be a bent function from F_2^{m-1} to F_2 , $\mathbf{a} \in F_2^{m-1}$ and M be an $(m - 1) \times (m - 1)$ nonsingular matrix. Define

$$g(\mathbf{x}^*) = f_0(\mathbf{x}) \oplus x_m f_0(\mathbf{x}) \oplus x_m \oplus x_m f_0(M\mathbf{x} \oplus \mathbf{a}),$$

where $\mathbf{x}^* \in F_2^m$. Then, $g(\mathbf{x}^*)$ is $(m - 1)$ -plateaued function. When M is identity matrix and $\mathbf{a} = 0$, then $g(\mathbf{x}^*) = f_0(\mathbf{x}) \oplus x_m$ is a particular plateaued function, which is called partially-bent function [10].

Example 3. [13] For odd $m \geq 3$, there exists Boolean function $f(\mathbf{x})$ on F_2^m with the form $f(\mathbf{x}) = g(\mathbf{x}) + \prod_{i=1}^{m-1} x_i$ and Walsh spectra belonging to

$$\{\pm 2^{(m+1)/2}, 0, (2^{(m+1)/2} - 4), \pm 4\},$$

where $\deg(g(\mathbf{x})) \leq m - 2$.

4 The Linear Span

In this subsection, we discuss the linear span of the sequences constructed from nonlinear functions over F_2^m . First, we recall some notations and results from [14]. Let $n = 2m$ and

$$Q_r = \{q \in Z_{2^n} \mid 0 < w(q) \leq r\},$$

where $w(q)$ denotes the hamming weight of the base-2 representation of q . It is easy to see that Q_r is the union of all the cyclotomic cosets mod $2^n - 1$ with hamming weight belonging to $[1, r]$. Let

$$\tilde{Q}_r = \{q \in Q_r \mid q \leq 2^j q \pmod{2^n - 1}, j = 1, \dots, n - 1\}$$

be the cyclotomic coset leaders of Q_r . Similarly, we define

$$E_r = \{q \in Q_r \mid w(q) = r, q = \sum_{i=1}^r 2^{v_i} \text{ with } v_{i_0} - v_{j_0} = m \text{ for some } i_0, j_0, 1 \leq i_0, j_0 \leq r\},$$

and

$$\tilde{E}_r = \{q \in E_r \mid q \leq 2^j q \pmod{2^n - 1}, j = 1, \dots, n - 1\}.$$

Assume that H is the nonlinear function of degree d in m variables given by

$$H(x_1, \dots, x_d) = \prod_{i=1}^d x_i.$$

Let $\{\beta, \beta^2, \dots, \beta^{2^{m-1}}\}$ is a normal basis for F_{2^m} over F_2 . Set $x_i = \text{tr}_1^n(\beta^{2^{(i-1)}}\alpha^t)$, $i = 1, \dots, d$. Then, the trace expansion of $H\{\text{tr}_1^n(\beta^{2^{(i-1)}}\alpha^t), i = 1, \dots, d\}$ has the following form

$$H\{\text{tr}_1^n(\beta^{2^{(i-1)}}\alpha^t), i = 1, \dots, d\} = \sum_{q \in \tilde{Q}_d} \text{tr}_1^{p_q}(a_q \alpha^{qt}).$$

Let $w(q_0) = d$ and $q_0 = \sum_{i=1}^d 2^{v_i}$, where the integers v_i are all distinct. Then

$$a_{q_0} = \det \begin{bmatrix} \beta^{e_1} & \beta^{e_2} & \dots & \beta^{e_d} \\ \beta^{2e_1} & \beta^{2e_2} & \dots & \beta^{2e_d} \\ \vdots & \vdots & \ddots & \vdots \\ \beta^{2^{d-1}e_1} & \beta^{2^{d-1}e_2} & \dots & \beta^{2^{d-1}e_d} \end{bmatrix},$$

where $e_i = 2^{v_i}$, $i = 1, 2, \dots, d$. Setting $\eta_i = \beta^{e_i}$, then

$$a_{q_0} = \det \begin{bmatrix} \eta_1 & \eta_2 & \dots & \eta_d \\ \eta_1^2 & \eta_2^2 & \dots & \eta_d^2 \\ \vdots & \vdots & \ddots & \vdots \\ \eta_1^{2^{d-1}} & \eta_2^{2^{d-1}} & \dots & \eta_d^{2^{d-1}} \end{bmatrix}. \tag{6}$$

Lemma 1. [3] Let $\eta_1, \eta_2, \dots, \eta_d$ be elements of F_{q^m} , where q is a power of a prime. Then,

$$\begin{aligned} & \det \begin{bmatrix} \eta_1 & \eta_2 & \dots & \eta_d \\ \eta_1^q & \eta_2^q & \dots & \eta_d^q \\ \vdots & \vdots & \ddots & \vdots \\ \eta_1^{q^{d-1}} & \eta_2^{q^{d-1}} & \dots & \eta_d^{q^{d-1}} \end{bmatrix} \\ &= \eta_1 \prod_{j=1}^{d-1} \prod_{c_1, \dots, c_j \in F_q} \left(\eta_{j+1} - \sum_{k=1}^j c_k \eta_k \right), \end{aligned}$$

and so the determinant is $\neq 0$ if and only if $\eta_1, \eta_2, \dots, \eta_d$ are linearly independent over F_q .

Applying Lemma 1 to Eq. (6), we have the following proposition 3, which plays an important role in determining the linear span of bent-function sequences in [14].

Proposition 3. [14] Let H be the nonlinear function of degree d in m variables given by

$$H(x_1, x_2, \dots, x_d) = \prod_{i=1}^d x_i$$

Let α be a primitive element of F_{2^n} and the set $\{\beta, \beta^2, \dots, \beta^{2^{m-1}}\}$ a normal basis for F_{2^m} over F_2 . Set $x_t = \text{tr}_1^n(\beta^{2^{i-1}}\alpha^t)$, $i = 1, 2, \dots, d$. In the trace expansion

$$H(x_i(t), i = 1, 2, \dots, m) = \sum_{q \in \tilde{Q}_d} \text{tr}_1^{p_q}(b_q \alpha^{qt}), \tag{7}$$

if $w(q_0) = d$ and $q_0 \notin \widetilde{E}_d$, then $b_{q_0} \neq 0$. Thus, there are at least

$$|(Q_d \setminus Q_{d-1}) \setminus E_d| = \binom{m}{d} 2^d$$

different terms with nonzero coefficient in the polynomial representation of Eq. (7).

Similar as Kumar’s discussion before Theorem 3 in [14], using proposition 3 we can prove the following theorem which is the generalization of Theorem 3 in [14].

Theorem 2. For any integer $m \geq 2$, there exists nonlinear function from F_2^m to F_2 of degree d such that the sequences $s_{\eta,\theta}(t)$ constructed in Eq. (5) satisfying the following properties:

(i) The maximal correlation $\leq 2^{m+1} + 1$;

(ii) Each sequence within \mathcal{S}_θ and the linear span $l(s_{\eta,\theta}(t))$ of the sequences $s_{\eta,\theta}(t)$ satisfies the lower bound

$$l(s_{\eta,\theta}(t)) \geq \begin{cases} \binom{m}{d} 2^d + \frac{1}{2} \sum_{i=2}^{d-1} \binom{m}{d} 2^i + n, & \text{for } m > 4, 2 < d \leq \lfloor m/2 \rfloor; \\ \binom{m}{2} 2^{d-1} + n, & \text{for } m \geq 4, d = 2; \\ \binom{m}{d} 2^d + n, & \text{for } 4 > m \geq 2, d = 2. \end{cases}$$

The lower bound given in Theorem 2 can only be applied to special nonlinear functions with degree less than $\lfloor m/2 \rfloor$. Next, we give a lower bound with no restriction on the degree.

Theorem 3. For any integer $m \geq 2$, let D is subset of $\{1, 2, \dots, m\}$ with $|D| = d$. $f(\mathbf{x}) = g(\mathbf{x}) + \prod_{i \in D} x_i$ is a Boolean function with $g(\mathbf{x})$ on F_2^m satisfying $\text{deg}(g(\mathbf{x})) < d$, then the linear span of the sequence $s_{\eta,\theta}(t)$ constructed in Eq. (5) is at least $\binom{m}{d} 2^d + n$.

Proof. Since the algebraic degree of $f(\mathbf{x})$ is d , thus the sequence $s_{\eta,\theta}(t)$ has the following expression

$$s_{\eta,\theta}(t) = \sum_{q \in \widetilde{Q}_d} \text{tr}_1^{p_q}(a_q \alpha^{qt}) + \text{tr}_1^n[(\eta + \theta\sigma)\alpha^t], \tag{8}$$

where the coefficients a_q belong to the subfield F_{2^m} as they are purely the functions of the elements $\beta_i, i = 1, 2, \dots, m$ [15]. Since $a_q \in F_{2^m}$ and $\eta + \theta\sigma \in F_{2^n} \setminus F_{2^m}$, thus the term $\text{tr}_1^n(\theta\sigma\alpha^t)$ can not disappear in Eq. (8) and each sequence in the family \mathcal{S}_θ has the same linear span. Considering the trace expansion of $s_{\eta,\theta}(t)$, the terms $\text{tr}_1^{p_q}(a_q \alpha^{qt})$ with $w(q) = d$ only come from the sole degree d term in the function $f(\mathbf{x})$. Without loss of generality, we assume that

$$\prod_{i \in D} x_i = H(x_1, \dots, x_d) = \prod_{i=1}^d x_i.$$

By proposition 3, we know that in the trace expansion of $s_{\eta,\theta}(t)$, there are at least $|(Q_d \setminus Q_{d-1}) \setminus E_d| = \binom{m}{d} 2^d + n$ different terms. Thus, the lower bound of the linear span of $s_{\eta,\theta}(t)$ is $\binom{m}{d} 2^d + n$. The proof is finished.

Remark. Using the nonlinear function in Example 3, the constructed sequences have linear span larger than $n(2^{n/2-2} + 1)$ and maximal correlation $2^{m+1} + 1$. Although this lower bound is not so tight as that in Theorem 2, it has no restriction on degree less than $\lfloor m/2 \rfloor$.

5 Conclusion

The construction method of bent-function sequences is generalized. Using this generalized method, we can construct families of sequences with low correlation and balance property from any nonlinear functions with low Walsh transformation. We proved an interesting result that the maximal correlation of the constructed sequence family \mathcal{S}_θ is only dependent on the maximum magnitude spectra of the nonlinear function. Some examples are also given to illustrate our general construction method. The linear span of these sequences is also determined.

Acknowledgments. The authors would like to thank the anonymous reviewers for their helpful comments. The authors research is supported by Hubei Key Laboratory of Applied Mathematics and Science Foundation of South-Central University for Nationalities.

References

1. Olsen, J.D., Scholtz, R.A., Welch, L.R.: Bent-Function Sequences. *IEEE Trans. Inform. Theory* 28, 858–864 (1982)
2. No, J.-S., Gil, G.-M., Shin, D.-J.: Generalized Construction of Binary Bent Sequences with Optimal Correlation Property. *IEEE Trans. Inform. Theory* 49, 1769–1780 (2003)
3. Lidl, R., Niederreiter, H.: *Finite Fields*. Addison-Wesley, Massachusetts (1983)
4. Khoo, K., Gong, G., Stinson, D.R.: A New Characterization of Semi-Bent and Bent Functions on Finite Fields. *Des., Codes, Cryptogr.* 38, 279–295 (2006)
5. Rothaus, O.S.: On Bent Functions. *J. Comb. Theory (Ser. A)* 20, 300–305 (1976)
6. Zheng, Y., Zhang, X.M.: Plateaued Functions. In: Varadharajan, V., Mu, Y. (eds.) *ICICS 1999*. LNCS, vol. 1726, pp. 284–300. Springer, Heidelberg (1999)
7. Zhang, X.M., Zheng, Y.: On Plateaued Functions. *IEEE Trans. Inform. Theory* 47, 1215–1223 (2001)
8. Chabaud, F., Vaudenay, S.: Links Between Differential and Linear Cryptanalysis. In: De Santis, A. (ed.) *EUROCRYPT 1994*. LNCS, vol. 950, pp. 356–365. Springer, Heidelberg (1995)
9. Dobbertin, H.: Almost Perfect Nonlinear Power Functions on $\text{GF}(2^n)$: The Welch Case. *IEEE Trans. Inform. Theory* 45, 1271–1275 (1999)

10. Carlet, C.: Partially-Bent Functions. *Designs, Codes and Cryptography* 3, 135–145 (1993)
11. Budaghyan, L., Carlet, C., Pott, A.: New Classes of Almost Bent and Almost Perfect Nonlinear Polynomials. *IEEE Trans. Inform. Theory* 52, 1141–1152 (2006)
12. Chee, S., Lee, S., Kim, K.: Semi-Bent Functions. In: Safavi-Naini, R., Pieprzyk, J.P. (eds.) *ASIACRYPT 1994*. LNCS, vol. 917, pp. 107–118. Springer, Heidelberg (1995)
13. Maitra, S., Sarkar, P.: Cryptographically Significant Boolean Functions with Five Valued Walsh Spectra. *Theoretical Computer Science* 276, 133–146 (2002)
14. Kumar, P.V., Scholtz, R.A.: Bounds on the Linear Span of Bent Sequences. *IEEE Trans. Inform. Theory* 29, 854–862 (1983)
15. Kumar, P.V.: *Analysis and Generalization of the Bent Function Sequences*. Ph.D. dissertation in electrical engineering, University of Southern California

An Efficient Large-Scale Volume Data Compression Algorithm

Degui Xiao¹, Liping Zhao^{1,2}, Lei Yang¹, Zhiyong Li¹, and Kenli Li¹

¹ School of Computer and Communication, Hunan University, Changsha 410082, China

² School of Information Engineering, Jiaying University, Jiaying 314000, China

Abstract. Considering empty region in the volumetric data occupying a certain percentage, an efficient large-scale data compression algorithm based on VQ is presented. Firstly, the entire volume data are divided into many smaller regular blocks, and the blocks are classified into two groups according to their average gradient values: one consists of those blocks with zero average gradient value, and the other consists of those with non-zero average gradient values. Secondly, only those blocks with non-zero average gradient values are decomposed into a three hierarchical representation and vector quantized. Finally, block data in different groups are reconstructed with different ways. When applying this algorithm to the volume data, all experimental results demonstrate the proposed algorithm is more efficient than most existing large-scale volume data compression algorithms.

Keywords: Vector quantization, Classifying, Volume data, Volume compression.

1 Introduction

Direct volume rendering of large volumetric data sets on programmable graphics hardware is often limited by the amount of available graphics memory and the bandwidth from main memory to graphics memory. Compressed Volume Rendering (CVR)[1] has been shown to be an effective solution to this problem. CVR is a general approach for combining volumetric compression and volume rendering such that the decompression is coupled to rendering. Vector quantization(VQ),with its relatively complex encoding and simple decoding that is essentially a single table look-up procedure, is an ideal choice for an asymmetric coding for CVR [1,2,3,4,5].Vector quantization has first been applied in volume rendering applications for compression purposed by [2]. [3] has developed a hierarchical VQ scheme (short for HVQ), which can reach good fidelity and however, about 3 times lower compression rate than the simple VQ (short for SVQ). [1] presented a novel volume compression method based on transform coding using the KLT and partitioned vector quantization. However, the compression algorithm is too complex and much more time consumed. To overcome the shortcomings of the relatively low compression rate of [3], this paper presents a new efficient large-scale data compression algorithm based on VQ, also called FCHVQ (Flag based Classical hierarchical VQ). The key to our approach are a subtle classification and flags set for each block during the data pre-process stage. Under the

premise of the good quality of image reconstruction, our proposed scheme can improve the compression rate significantly. Volume data have the same characteristic that they are not chaotic and usually empty region in the volumetric data occupies a certain percentage. Especially, when the block size is not too large, data in one block usually are highly related. Applying FCHVQ to the volume data obtains good performance.

2 Related Works

Vector quantization (VQ) maps every k -dimensional input vector X to some reproduction vector x_i selected from a finite codebook of M candidate vectors (code words), and encodes X by the index i of x_i . If M is small, then each index requires few bits ($\log_2 M$) and compression is achieved. In this paper, data compression rate can be evaluated by (1).

$$Rate_{comp} = \frac{OriginalDataSize}{CompressedDataSize} \quad (1)$$

2.1 Simple VQ

Simple VQ uses a conventional vector quantizer. Comparing to other VQ algorithms, it can achieve higher compression rate, however, with bad image reconstruction quality. Consider a volume of size $N \times M \times K$ each of whose point holds B bytes, and suppose the block size is $n \times n \times n$ and the capacity of the codebook is $N_{codebook}$, then compression rate of SVQ is defined by (2).

$$SVQ_Rate_{comp} = \frac{N \times M \times K \times B}{N \times M \times K / (n \times n \times n) + N_{codebook} \times n \times n \times n} \quad (2)$$

For simplicity, we assume that the data bits after compression is 8, and the number of code words in the codebook is 256, so the bits of each index is also 8 ($\log_2 256 = 8$). On the other hand, for volume rendering, data should be usually normalized between 0 to 255 which can also be presented within 8 bits before applying the transform function.

2.2 Hierarchical VQ

Hierarchical VQ was described by [3] to get better image reconstruction quality. Each block is decomposed into a multi-resolution representation similar to wavelet manner. Specifically, starting with the original scalar field, the data is initially partitioned into disjoint blocks of size 4^3 . Each block is decomposed into a multi-resolution representation, which essentially splits the data into three different triadic frequency bands. Therefore, each block is down-sampled by a factor of two by averaging disjoint sets of 2^3 voxels each. The difference between the original data samples and the respective down-sampled value is stored in a 64-component vector. The same process is applied to the down-sampled version, producing one single value that represents the mean value of the entire block. The 2^3 difference values carrying the information that

is lost when going from 2^3 mean values to the final one are stored in an 8-component vector. Finally, a 1-component vector stores the mean of the entire block. In performing this task, the data is decomposed into three vectors of length 64, 8, and 1, respectively, which hierarchically encode the data samples in one block. In this way, HVQ can reach good fidelity, however, much lower compression rate because each block should store the mean value of the block and two indices in order to reconstruct the whole block value. Also, consider a volume with size $N \times M \times K$ each of whose point holds B bytes, and the block size is $n \times n \times n$ and the down-sample block size is $(n/2) \times (n/2) \times n/2$, then compression rate of SVQ can be computed by (3).

$$HVQ_Rate_{comp} = \frac{N \times M \times K \times B}{3 \times N \times M \times K / (n \times n \times n) + C_1 \times n \times n \times n + C_2 \times n \times n \times n / 8} \tag{3}$$

Lemma 1. When the block size used in HVQ is same as that used in SVQ, $SVQ_Rate_{comp} / HVQ_Rate_{comp} \approx 3$.

Proof. Just as mentioned before, the main impact of the compression rate is the storage of index value. When the volume data size is big enough, the codebook capacity can be ignored, then

$$SVQ_Rate_{comp} \approx \frac{N \times M \times K \times B}{N \times M \times K / (n \times n \times n)}$$

And,
$$HVQ_Rate_{comp} = \frac{N \times M \times K \times B}{3 \times N \times M \times K / (n \times n \times n)}$$

So,
$$SVQ_Rate_{comp} / HVQ_Rate_{comp} \approx 3$$

3 Flag Based Classical Hierarchical VQ

The study of VQ is mainly about how to improve the compression rate and reduce distortion and algorithm complexity [2]. However, in most existing algorithms, the increment of the compression rates tends to lead to lower quality of image reconstruction (such as SVQ), and improvement of the quality of image reconstruction leads to lower compression rate (such as HVQ).

In order to not only obtain a good reconstruction quality but also improve the compression rate, this paper presents a new efficient large-scale data compression algorithm based on VQ, also called FCHVQ. The overall algorithm is illustrated in Fig.1. During the data pre-process stage, classify the blocks which divided from total volume data according to average gradient values into two groups: one containing blocks with zero average gradient values and the other containing blocks with non-zero average gradient values, and then set up a flag for each block. Only those blocks with non-zero average gradient values are decomposed into a three hierarchical representation and vector quantized similar to [3]. Blocks with zero average gradient values just store the mean value of each block. In the decode stage, reconstruct each block in different way as the flag varies.

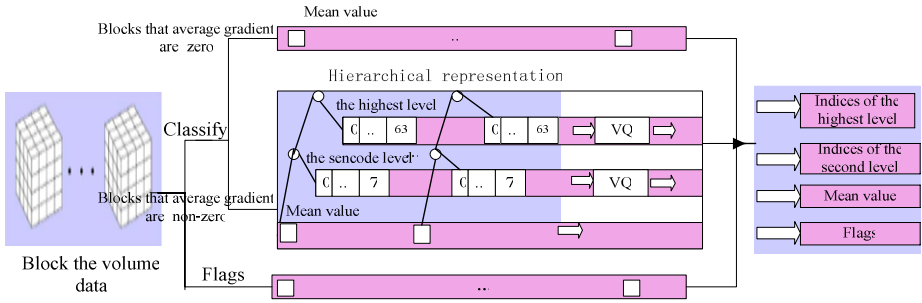


Fig. 1. FCHVQ algorithm is illustrated

3.1 Data Pre-Process of FCHVQ

FCHVQ, fully considering the volume itself characteristics, firstly computes the gradient similarly with the way of forward-difference, secondly classifies the blocks according to their average gradient values. Suppose $H(x, y, z)$ is the gradient value of X , and $f(x, y, z)$ is the original data point value at position (x, y, z) , then

$$H(x, y, z) = |f(x+1, y, z) - f(x, y, z)| + |f(x, y+1, z) - f(x, y, z)| + |f(x, y, z+1) - f(x, y, z)| \tag{4}$$

$AvH(B_i)$ is the average gradient value of each block, it can be defined by (5):

$$AvH(B_i) = \frac{1}{n \times n \times n} \sum_{x=x_i}^{x_i+n} \sum_{y=y_i}^{y_i+n} \sum_{z=z_i}^{z_i+n} H(x, y, z) \tag{5}$$

From (4) and (5), we can clearly see that if the data in one block have the same value, $AvH(B_i)$ equals zero. And if the data in one block varies little, the value of $AvH(B_i)$ is relatively very small. Taking into account that the volume data is not chaotic, the data in the block are possibly same. Many experiments show that the gradient values of non-zero blocks always occupy a certain proportion of the total blocks. By applying a simple threshold, if $AvH(B_i)$ of a block is less than the threshold, the block is classified into a group containing blocks with zero average gradient value, otherwise into another group containing blocks with the non-zero average gradient values. Fig.2 illustrates the FCHVQ data pre-process algorithm.

Considering a volume has $N \times M \times K$ data points and each point holds B bytes, the block size is $n \times n \times n$ and the down-sample block size is $(n/2) \times (n/2) \times n/2$, the compression rate of FCHVQ can be computed by (6).

$$FCHVQ_Rate_{comp} = \frac{N \times M \times K \times B}{9 \times N \times M \times K / (2n \times 2n \times 2n) + 2 \times B_{g=0} + C_1 \times n \times n \times n + C_2 \times n \times n \times n / 8} \tag{6}$$

Where, $B_{g=0}$ is the number of blocks holding non-zero gradient values, $C_1 \times n \times n \times n$ represents the capacity of the codebook in the highest hierarchical level,

and the $C_2 \times n \times n \times n / 8$ represents the capacity of the codebook in the second hierarchical level. Because we only save the indices of non-zero blocks, so $2 \times B_{g=0}$ would be occupied. Finally the mean value and the flag of each block are $N \times M \times K / (n \times n \times n)$ and $N \times M \times K / (n \times n \times n \times 8)$, respectively. And their summation is $9 \times N \times M \times K / (2n \times 2n \times 2n)$.

```

encode(raw_data, rx, ry, rz, compression_data)
Begin
    rx4=(rx>>2), ry4=(ry>>2), rz4=(rz>>2);
    totalBlocks = rx4* ry4* rz4;
    BlockVolData(raw_data, 4, rx4, ry4, rz4);
    for i=0 to totalBlocks do
        begin
            getAverage( block[i] );
            getGradient( block[i] );
        end;
    for i=0 to totalBlocks do
        begin
            if( block[i].gradient <= threshold) then
                begin
                    block[i].group=0;
                    compression_data->flag[i]=0;
                    class0++;
                end;
            else
                begin
                    block[i].group =1;
                    compression_data->flag[i]=1;
                    class1++;
                end;
            end;
        end;
    /*Gradient value of non-zero blocks are decomposed into
    a three hierarchical representation manner like [3];*/
    .....
end;

```

Fig. 2. FCHVQ data pre-process algorithm

Lemma 2. Suppose $B_{g=0}$ is the number of blocks with zero gradient values and when $B_{g=0}$ is more than 6.25 percent of the total number of blocks, in other words, $B_{g=0} / (B_{g=0} + B_{g \neq 0}) > 1/16$, then $FCHVQ_Rate_{comp} > HVQ_Rate_{comp}$.

Proof. If $FCHVQ_Rate_{comp} > HVQ_Rate_{comp}$, then

$$9 \times N \times M \times K / (2n \times 2n \times 2n) + 2B_{g \neq 0} \text{ should be less than } 3 \times N \times M \times K / (n \times n \times n).$$

$$\text{Then } B_{g \neq 0} < 15 \times (N \times M \times K) / 16,$$

add $B_{g=0}$ on both sides of the equation, get:

$$B_{g=0} + B_{g \neq 0} < B_{g=0} + 15 \times N \times M \times K / 16,$$

$$\text{Because } (B_{g=0} + B_{g \neq 0}) = N \times M \times K / n \times n \times n,$$

$$\text{So } B_{g=0} > N \times M \times K / (16 \times n \times n \times n),$$

divided by $B_{g=0} + B_{g \neq 0}$ on both sides of the equation, get:

$$B_{g=0} / (B_{g=0} + B_{g \neq 0}) > 1/16.$$

$$\text{So, if } B_{g=0} / (B_{g=0} + B_{g \neq 0}) > 1/16,$$

$$\text{then } FCHVQ_Rate_{comp} < HVQ_Rate_{comp}.$$

Considering empty region in the volumetric data occupies a certain percentage, many experiments show that the number of blocks whose average gradient values are zero is much more than 6.25 percent of the total number of blocks, that's the key point why our algorithm can be efficient. On the other hand, when the number of blocks whose average gradient values are zero is less than 6.25 percent of the total number of blocks, we can study how to divide the data into blocks in order to get more number of blocks whose average gradient values are zero.

3.2 Encoding of FCHVQ

Blocks with non-zero average gradient values are decomposed into a three hierarchical representation and should be vector quantized. The process of VQ can be divided into three aspects: code book design, encoding and decoding [6]. Code book design plays an important role in the performance of the algorithm. [10] developed one of the earliest vector quantization algorithms suitable for practical applications, the LBG-algorithm. Because of the sensitivity to the initial codebook in LBG, so far many optimized algorithms [7][8][9][11] have been proposed. Similar to [3], we use splitting scheme based on a principal component analysis (PCA) to find an initial codebook which is then refined by LBG algorithm, and restrict the search to the k-neighborhood of the initial cell in the quantization stage. Not like method in [3], in our method, only those blocks with non-zero average gradient values are trained. Thus, we do not only save a large amount of computation, but also leave the whole code words to the blocks whose average gradient values are non-zero.

3.3 Decoding of FCHVQ

The main idea of decoding algorithm of FCHVQ is to reconstruct the whole data in each block according to the saved information. Firstly, we obtain the beginning and the ending indices in all three directions to determine the position of the block in the volume data. Then look into the flag of each block. If the flag is zero, we can reconstruct the whole block with its mean value. Otherwise, we should first get its mean

value V_{mean} , then get the difference V_{D1} between V_{mean} and value in the second down-sample block according to the second level index, finally get the difference V_{D2} between the first down-sample block and that in the second down-sample block according to the highest level index. At last we use the sum of V_{mean} , V_{D1} and V_{D2} as the reconstruction value of the block.

Different from the decoding algorithm of HVQ, for those blocks with average zero gradient values, we just replace their whole block data with their mean values. Evidently, our method is faster than HVQ. Experiment shows that, for all the testing data, FCHVQ algorithm decodes faster than HVQ algorithm.

3.4 Results and Comparison

In order to provide a context for the evaluation of our work, we compare our approach (FCHVQ) with analogous implementations of SVQ and HVQ.

The performance of VQ is measured by the compression rate and the reconstructed image quality. The reconstructed image quality is evaluated by the peak signal to noise ratio (PSNR) [6]. Here, the number of codeword in the codebook is 256. The size of volume data bonsai, aneurism and foot is $256 \times 256 \times 256 \times 8$ bits. We also extend our method to the seismic volume data compression [12][13][14]. The size of seismic data is $1024 \times 256 \times 256 \times 16$ bits. The block size is 64.

Table 1. Comparison between SVQ, HVQ and FCHVQ

Volume data	HVQ		FCHVQ		SVQ	
	PSNR (db)	Compression rate	PSNR (db)	Compression rate	PSNR (db)	Compression rate
bonsai	36.27	20.84	36.36	36.16	26.63	60.24
aneurism	36.21	20.84	36.26	51.08	29.45	60.24
foot	32.28	20.84	32.39	35.16	21.73	60.24
seismic	30.40	42.42	30.79	53.00	23.80	126.0

From table 1, we can see that FCHVQ can get higher compression rate than HVQ in the premise of better reconstruction image quality than SVQ.

For the volume data aneurism, table 2 shows the comparison between SVQ, HVQ, FCHVQ when different codebook sizes, 64, 128 and 256 are used, from which we can see that FCHVQ is more efficient than others even in different codebook size.

Table 2. Comparison between different codebook sizes of the HVQ, FCHVQ, SVQ

Code book size	HVQ		FCHVQ		SVQ	
	PSNR (db)	Compression rate	PSNR (db)	Compression rate	PSNR (db)	Compression rate
64	34.55	21.21	34.58	53.33	28.30	63.02
128	35.34	21.09	35.36	52.56	29.13	62.06
256	36.21	20.84	36.26	51.08	29.45	60.24

In the CVR domain, decompression is coupled to rendering. Compression can be slow since it is performed only once offline. But the decompression should be extremely fast since it is performed many times per second during rendering. FCHVQ, which needs compute the average gradient of each block, really consumes a little more time during compression than HVQ. However, when decoding, because FCHVQ can replace the whole block data with the mean value of the block if the average gradient is zero, it runs faster than HVQ. Take Aneurism for example, HVQ needs 6.17 seconds while FCHVQ only needs 3.31s to decompress.

Fig.3 is a comparison of the original volume data and the reconstruction of compressed volume data by FCHVQ (visualized by Open Inventor). These following sequences demonstrate the effectiveness of FCHVQ.

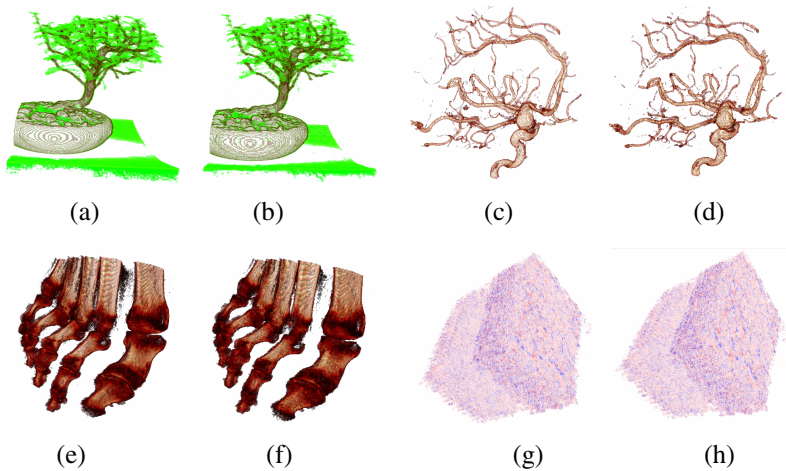


Fig.3. Comparison of visualization results between the original and the compressed volume data. (a) is the original “bonsai” and (b) is the compressed “bonsai”. (c) is the original “aneurism” and (d) is the compressed “aneurism”. (e) is the original “foot” and (f) is the compressed “foot”. (e) is the original “seismic” and (f) is the compressed “seismic”.

4 Conclusion and Future Work

In order to relieve the conflict between the compression rate and reconstructed image quality, this paper presents an efficient large-scale data compression algorithm based on VQ. The key to our algorithm is to give full consideration of the characteristics of the volume data itself. Noticing that the data in a block may has the same value when the block size is not too big, we present a subtle classification scheme before VQ. While applying proposed algorithm, FCHVQ, to the seismic data and other testing data sets, the experimental results show that this algorithm can not only obtain a better image reconstruction quality, but also increase the compression rate significantly. In the future, we will investigate how to use our scheme to decompress and do rendering on GPU.

Acknowledgements

This work is supported by the National Natural Science Foundation of China under Grant Nos.90715029.

References

1. Fout, N., Ma, K.L.: Transform Coding for Hardware-Accelerated Volume Rendering. *J. IEEE Visualization and Computer Graphics* 13, 1600–1607 (2007)
2. Ning, P., Hesselink, L.: Fast Volume Rendering of Compressed Data. In: *Workshop Volume Visualization*, pp. 11–18. IEEE Press, San Jose (1993)
3. Schneider, J., Westermann, R.: Compression Domain Volume Rendering. In: *Visualization 2003*, pp. 239–300. IEEE Press, Seattle (2003)
4. Fout, N., Ma, K.L., Ahrens, J.: Time-Varying, Multivariate Volume Data Reduction. In: *ACM Symposium on Applied Computing*, pp. 1224–1230 (2005)
5. Guo, d., Cheng, Q.S., Sun, X.C.: Vector Quantization Based Shear-Warped Volume Rendering. *J. Computer Aided Design & Computer Graphics* 13, 532–536 (2001) (in Chinese)
6. Sun, S.H., Lu, Z.M.: *Technology and Application of Vector Quantization*. Science Press, Beijing (2002) (in Chinese)
7. Sun, H.W., Dong, W.M., Song, B.H.: Codebook Design Algorithm Based on Principal Component Analysis and Genetic Algorithm. *J. Computer Aided Design & Computer Graphics* 16, 1651–1655 (2004) (in Chinese)
8. Lu, Z.M., Pan, Z.X., Sun, S.H.: VQ Codebook Design Based on the Modified Tabu Search Algorithms. *J. Acta Electronica Sinica* 28, 108–110 (2000) (in Chinese)
9. Equitz, W.: A New Vector Quantization Clustering Algorithm. *J. IEEE Transactions on Acoustics, Speech and Signal Processing*, 1568–1575 (1989)
10. Linde, Y., Buzo, A., Gray, R.: An Algorithm for Vector Quantizer Design. *J. IEEE Transactions on Communications* 28, 84–95 (1980)
11. Liao, H.L., Ji, Z., Wu, Q.H.: A Novel Genetic Particle-Pair Optimizer for Vector Quantization in Image Coding. In: *IEEE World Congress on Computational Intelligence*, pp. 708–713. IEEE Press, Hong Kong (2008)
12. Averbuch, A.Z., Meyer, R., Stromberg, J.O., Coifman, R., Vassiliou, A.: Low Bit-Rate Efficient Compression for Seismic Data. *J. IEEE Transactions on Image Processing* 10, 1801–1814 (2001)
13. Duval, L.C., Nagai, T.: Seismic Data Compression Using GULLOTS. *J. IEEE Transactions on Signal Processing* 49, 1765–1768 (2001)
14. Yilmaz, O., Huang, X.D., Yuan, M.D.: *Translate Seismic Data Processing*. Petroleum Industry Press, Beijing (1994) (in Chinese)

Simultaneous Synchronization of Text and Speech for Broadcast News Subtitling

Jie Gao, Qingwei Zhao, Ta Li, and Yonghong Yan

ThinkIT Speech Lab, Institute of Acoustics, Chinese Academy of Sciences,
Beijing 100190, P.R. China

{jgao,qwzhao,tli,yonghong.yan}@hcc1.ioa.ac.cn

Abstract. In this paper, we present our initial effort in automatic generation of subtitle for live broadcast news programs, utilizing the fact that nearly perfect transcriptions are available. Instead of using the former error-prone automatic-speech-recognition (ASR)-based method, we propose to formulate the subtitling problem as synchronization of text and speech, which is further simplified into an anchor points estimation problem. The Viterbi algorithm for hidden Markov model (HMM) is augmented with new criterions for the online anchor points estimation. Experiments indicate that our proposed methods show satisfying performance for the simultaneous subtitling application. We also present a brief introduction into our whole subtitling system under further development.

Keywords: Live broadcast news subtitling, Text and speech synchronization, Hidden Markov model.

1 Introduction

There is a generic request from society and governments that are pushing the TV broadcasters to increase the amount and diversity of programs subtitled. While TV programs with subtitles become increasingly available in US, Canada, Japan, French and Portugal [1,2], the ratio of subtitled TV programs in China is still low. For the live programs, such as broadcast news, the ratio is even lower.

Most of the programs with subtitles are prerecorded, where the subtitles are created by manual transcription of the speech and align them to the spoken contents manually or automatically [3,4,5]. Typically, in automatic alignment approaches, to deal with the erroneous transcriptions and varied acoustic condition, the alignment is actually converted to an automatic speech recognition (ASR) problem. Time-marked transcriptions are first produced by an ASR system, then they are aligned to the reference text to yield segmentation points of sentences. Because these approaches essentially work in an offline mode, which require the whole speech content before processing, only programs produced prior to its broadcasting day can be subtitled.

In order to implement simultaneous subtitling for live programs, particularly news programs, simultaneous captions are created manually by skilled people called *Stenographer* [1] in US and Canada. However, according to [2] the manual

captioning of live speech in Chinese/Japanese is much more difficult because they contain many characters of homonyms, and the stenographer has to select from multiple alternatives. Therefore, automatic methods are developed [1,2,6]. To my knowledge, the developed automatic techniques for live programs are essentially all based on automatic speech recognition (ASR). Although the typical word error rate (WER) is about 5% exploiting the prior information about the transcriptions, instant manual error correction is still needed. This causes a lag of 5 seconds or more in subtitle display compared to corresponding speech contents [2].

This paper constraints the subtitling of live TV programs with nearly perfect electronic transcriptions available. This assumption is true for the most of news broadcasting, where the reported contents are carefully checked and edited before the announcers finally read them on-air. In this scenario, the subtitling of live programs is formulated as a problem of automatic simultaneous synchronization (online alignment) between speech and text. Specifically, given a nearly perfect transcription of a news item as a set of textual sentences, we seek the begin time of each sentence and the display the sentence when the announcer begins to read it out; we seek the end time when the announcer finishes reading it and erase the text from the TV screen.

Although automatic aligning speech to corresponding text might seem a trivial problem [7,8] within the hidden Markov model (HMM)-based framework, few works are reported to be used for subtitling for live programs. The major reason may lie in the subtitling operation for live programs implies, besides real-time, an online operation. Transforming existing algorithms to online operation presents several problems. For instance, the first problem is the previous work via Viterbi algorithm with HMM essentially runs in an offline mode, which requires traceback to make the alignment decision. In addition, although we constraint ourselves to cases where nearly perfect transcriptions are available, alignment errors are still unavoidable in case of speech segments with significant background noise, casually spoken speech and even field reports without corresponding transcriptions. Any error may be fatal. Because the synchronization between online speech and text proceeds in a sequential manner and any error may make the subsequent speech lose synchronization with its text. Therefore we try to overcome these difficulties and build a realistic subtitling system. In this paper, we give an overview of our system under development and focus on our achievement on solution of the first problem.

This paper is organized as follows. We briefly introduce our subtitling system under development in Section 2. Then we detail our approach to synchronize the text and speech in Section 3. The system's performance is evaluated in Section 4. Section 5 concludes the comments on current and future work.

2 System Overview

Our simultaneous broadcast news subtitling system consists of the following modules: the text processor, the text-speech synchronizer, the error detector and error corrector. Fig.1 presents the system architecture on a conceptual level.

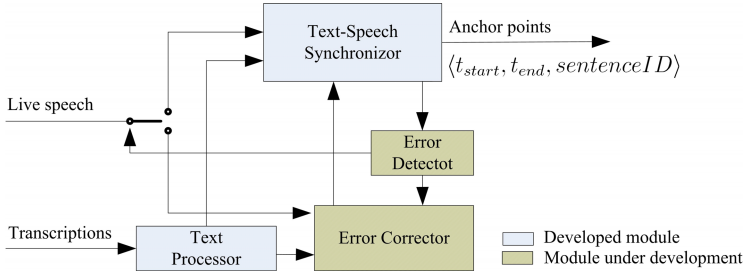


Fig. 1. Overall architecture of simultaneous broadcast news subtitling system

The text processor preprocesses the transcription of news items before the announcers read them out on-air. It has two main roles: 1) It normalizes the text and converts the out-of-vocabulary words (especially numbers) into in-vocabulary words. 2) It segments the text stream into a set of sentences that are appropriate for display on TV screen. Since its implementation is trivial, it will be ignored in the following of this paper.

The text-speech synchronizer is the key module. It explores methods devised in this work to detect the sentence start time and sentence end time (anchor points) in real time so that the texts can be displayed simultaneously with announcers’ speech. It is in essence an extension of conventional Viterbi-based aligner module. The workings of this component will be further detailed in the following sections.

The error detector module detects failure of text-speech synchronizer, which may be caused by missing transcription, adverse acoustic condition, etc. Whenever some error occurs, the error corrector, which consists of a speech recognizer followed by a text aligner, wait until speech segments which matches corresponding text well, and give feedback to the text-speech synchronizer, forcing it to work again. This module of is still under development.

3 Text-Speech Synchronization

As stated in last section, the key of the text-speech synchronization is the generation of the anchor points (sentence start and sentence end time corresponding to the speech). We formally formulate this problem by first defining the following notations.

- W_1^N : The transcription for a news item represented in a word sequence
- X_1^T : The speech corresponding to W_1^N represented in a feature vector
- S_1^M : A set of M sentences by segmenting W_1^N into $S_i^M = \{W_{n_{s_1}}^{n_{e_1}} \dots W_{n_{s_i}}^{n_{e_i}} \dots W_{n_{s_M}}^{n_{e_M}}\}$. We denote the i^{th} sentence as $S_i = \{W_{n_{s_i}}^{n_{e_i}}\}$, which is a subsequence of the W_1^N .
- $\{t_{s_i}, t_{e_i}, S_i\}$: Triple of start time, end time, sentence of the i^{th} sentence S_i .
- $\tau(W_n)$: Word end time of the n^{th} word W_n .

If some approximation are allowed, we can use the end time of the first word in the sentence as the sentence start time and use the time of the last word in the sentence as the sentence end time. Formally, we have

$$t_{s_i} \approx \tau(W_{n_{s_1}}) \tag{1}$$

$$t_{e_i} \approx \tau(W_{n_{e_1}}) \tag{2}$$

Although this approximation will introduce a lag of about a character long in display at the begin of a sentence, it's still feasible for the subtitling application because a lag of character is perceptually negligible. Therefore the problem are simplified to the problem of estimation the optimal word end time given $\tau(W_n)$ an *anchor word* W_n and all the speech input X_1^T . The optimality of the estimation for word end depends a specific criterion function $J(\tau|W_n, X_1^T)$ used:

$$\hat{\tau}(W_n) = \arg \max_{\tau} J(\tau|W_n, X_1^T), 1 < \tau < T \tag{3}$$

where $J(\tau|W_n, X_1^T)$ is the criterion function. In this paper, we denote this problem specified by (3), that is, to estimate the word end time of a *anchor word* an *anchor points estimation* problem.

3.1 Anchor Points Estimation Using Viterbi Criterion

We briefly overview the traditional speech-text alignment problem using HMM in an anchor point estimation perspective. Speech-text alignment via HMM is a simplified speech recognition problem. Given a word sequence W_1^N and pre-trained HMMs ϕ , a state-level search space is constructed by concatenating HMM models of the phoneme sequence corresponding to W_1^N specified by a pronunciation lexicon. The a state sequence s_1^K corresponding to W_1^N is obtained. The search space is essentially a state T-by-K trellis, as shown in Fig.2¹. And the alignment is to find the best path (state sequence) \hat{q}_1^T in the search space which maximize the joint probability of state sequence and the acoustic observation X_1^t :

$$\hat{q}_1^T = \max_{q_1^T} Pr(X_1^T, q_1^T|W_1^N), q \in \{s_1, \dots, s_m\} \tag{4}$$

which can be efficiently solved by the Viterbi algorithm [8,9,10]. When the whole alignment is finished, end time of each state s_k , can be obtained by the backtrace of the optimal Viterbi path. Optimal estimation of end time of a word (W_n) in Viterbi sense $\hat{\tau}_v(W_n)$ can be obtained as the end time of its word end state s_{we}

$$\hat{\tau}_v(W_n) = \arg \max_{\tau} J_V(\tau|W_n, X_1^T) \tag{5}$$

¹ This is a *conceptual* simplification, where the search space may be more complicated considering optional inter-word silences and multiple pronunciations of each word in practice.

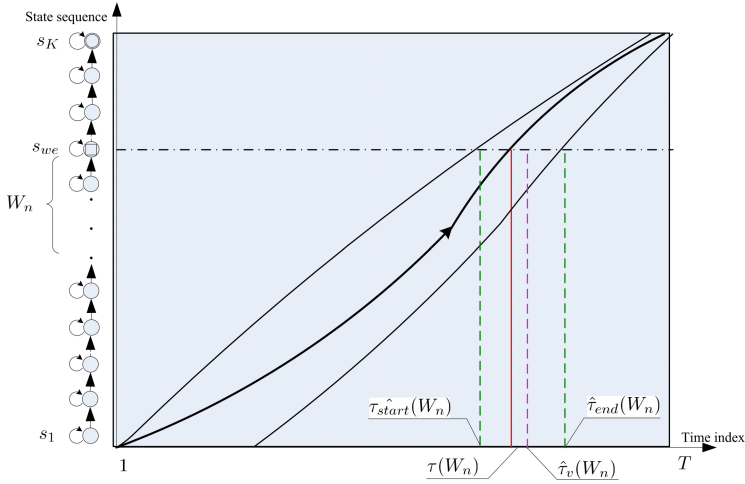


Fig. 2. Search space in the text-speech alignment

where $J_V(t|W_n, X_1^T)$ is the Viterbi criterion function specified by (4). Note that the Viterbi criterion function is globally optimal function and relies on all the speech data X_1^T , definitely including the data after time instant $\tau(W_n)$. Therefore, Viterbi alignment for anchor points estimation work in an offline mode and show good performance if no online operation is desired, that is when all X_1^T is known beforehand.

However, in the subtitling problem for live program, speech is input online. Therefore, estimation the anchor points must be made locally *around* its the real value so that the text can display simultaneously with the speech, which requires alternative estimation criterion functions.

3.2 Proposed Anchor Points Estimation Criteria

In practice, the full search is computationally very expensive when the state sequence is too long and only the most promising hypotheses are retained. In our system, the alignment is performed with a state-of-the-art LVCSR decoder, in which time-synchronous beam search is adopted [11]. Due to effective pruning, only part of the search space is probed, which contains the Viterbi-Optimal path, as shown in Fig.2. A side effect of beam pruning is that it actually constrains the estimation the anchor point $\tau(\hat{W}_n)$ within a range $[\hat{\tau}_{start}(W_n), \hat{\tau}_{end}(W_n)]$, as shown in Fig.2, which provide the basis for our proposed estimation criterions.

$$\hat{\tau}(W_n) = \arg \max_{\tau} J(\tau|W_n, X_1^T) \quad \hat{\tau}_{start}(W_n) < \tau < \hat{\tau}_{end}(W_n) \quad (6)$$

Therefore some heuristic driven criterion function are defined and deployed in the Viterbi alignment framework, inspired by the confidence estimation in keyword spotting domain [12].

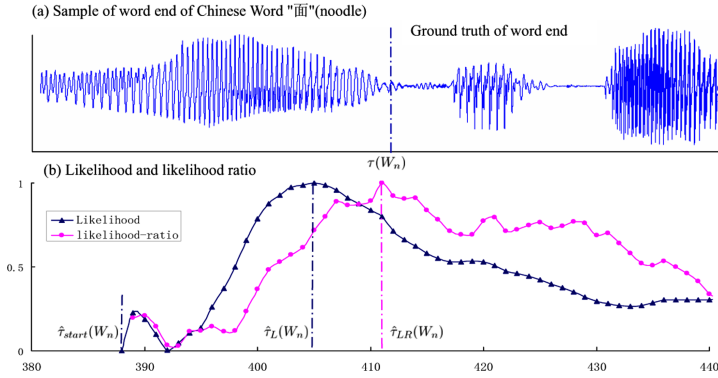


Fig. 3. Proposed likelihood/likelihood-ratio criterion around the true anchor point

- Top-Likelihood Criterion (TLC) The criterion function is the duration normalized likelihood of the ending word;

$$J_{TL}(\tau) = \frac{h(\tau, s_{we}, W_n)}{t_e - t_s}; \tag{7}$$

where s_{we} is the word end state of W_n ; $h(\tau, s_{we}, W_n)$ is the likelihood of the top-score word hypothesis of W_n given by the path hypothesis at time instant τ end on s_{we} ; t_s and t_e is the start time and the end time of the word hypothesis respectively.

- Top-Likelihood-Ratio Criterion (TLRC)

$$J_{TLR}(\tau) = \frac{\delta(\tau, s_{we})}{\sum \delta(\tau, s_{other})}; \tag{8}$$

where $\delta(\tau, s_{we})$ the path hypothesis landing on the word end state s_{we} of W_n at time instant τ ; $\delta(\tau, s_{other})$ the path hypothesis landing on the other state τ .

Fig.3 plots likelihood/likelihood-ratio around the ground-truth anchor point (word end time) as it changes over the time span constrained by the beam search. It shows the likelihood/likelihood-ratio reach a local peak around the ground-truth anchor point. We thus use the global maximum in likelihood/likelihood-ratio curve as the estimation of the anchor points. We have to remark that although location of the global optimal of likelihood/likelihood-ratio curve may require few frames of speech after the time instant of the maximum, it is still a local criterion compared to the Viterbi criterion reviewed in last section. Another point need to be pointed out is the proposed criterion functions only have to be calculated when with in the range constrained by the beam search, specifically when some path hypothesis end on the last state of the word. The are well suited for an online operation.

Once we can determine the anchor time points of the first word and last word of a sentence, the synchronization problem is solved as stated in (1)-(2). The text can be displayed simultaneously with the speech of the announcer.

4 Evaluation

In this section, we obtain objective measurement of the performance of our simultaneous subtitling system.

4.1 Data Corpora

News programs from Beijing Television Station (BTV) are used to evaluate our system. 24 news items and 430 sentences in total are randomly chosen as the evaluation speech. The ground-true anchor points are obtained via a two stage-procedure. First manual segmentation on news item are performed to segment the whole long audio into sentences; then text are aligned within each segment to obtain the ground-true of the word end. These news items consist of 430 sentences in total, which are further divided into three categories according to different acoustic conditions. 313 sentences are spoken by announcers with studio quality; 103 are by announcers, but with background noise. The rest 14 are sentences spoken by the reporters in the field reports or by the interviewees. They are summarized on Table 1.

Table 1. Categorization of the speech for evaluation

Categories	Acoustic Condition	# of sentences
Clean Speech	Announcers, studio quality	313
Noisy Speech	Announcers, background noise	103
Interview	Field reports, interviews	14

4.2 Results

We follow an evaluation metric in the past audio alignment work [13], i.e. the means and deviation of the error in estimating the sentence start points and sentence end points, given in seconds. We show the alignment error in Fig.4, for the both proposed criterion TLC and TLRC.

In analyzing the results from Fig.4, we see that the the proposed TLC and TLRC criterion show comparable performances and the TLRC is slightly better. Another point easily observed is that the starting points estimation are harder than the ending points (bigger mean in error and high standard derivation). This can be explained by the fact that the sentence ending points are constrained by the search process while the sentence begin is subjected to more variation. However, the system shows satisfying performance for simultaneous broadcast

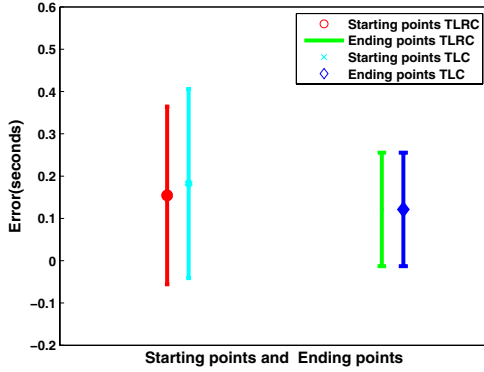


Fig. 4. System performance using proposed criteria

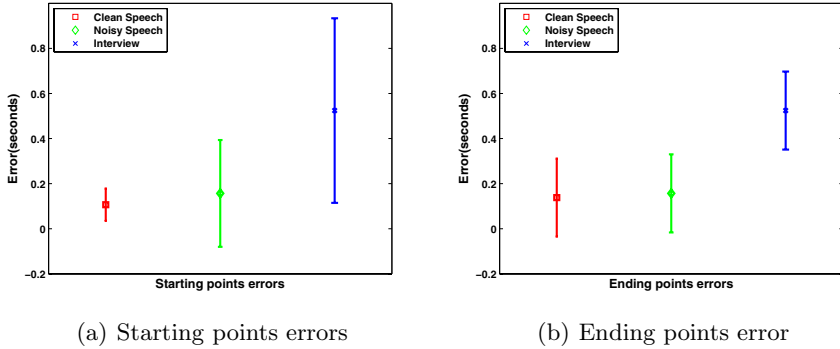


Fig. 5. System performance under different conditions

new subtitling, an average boundary error less than 0.2 seconds is perceptually negligible for a news audience.

In order to gain insight into performance our system, we reanalyze our systems’ performance under different conditions specified in Table 1, as shown in Fig.5. As expected, the performance degrades as the acoustic condition gets worse. The error in the interview part contributes to system error most and may affect the performance of clean speech around them. This necessitates development of the error detection/correction module as stated in Section 2.

5 Conclusion and Future Work

In this paper, we present our initial effort in automatic generation of subtitle for broadcast news programs where transcription are available. Instead of using the former computationally expensive and error-prone ASR-based method,

we propose to formulate the subtitling problem as synchronization of text and speech, which is further simplified into an anchor points estimation problem. New method based on likelihood/likelihood ratio are proposed for the online anchor points estimation. Experiments indicate our proposed method show satisfying performance for the simultaneous subtitling application. For future work, we will continue with our effort in system development, especially development of the error detection/correction module.

Acknowledgments. This work is partially supported by MOST (973 program, 2004CB318106), the National Natural Science Foundation of China (10574140, 60535030), the National High Technology Research and Development Program of China (863 program, 2006AA01010, 2006AA01Z195). Thanks for Changliang Liu and Haipeng Wang for fruitful discussion on the algorithm development.

References

1. Neto, J., Meinedo, H., Viveiros, M., Cassaca, R., Martins, C., Caseiro, D.: Broadcast News Subtitling System in Portuguese. In: ICASSP 2008: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1561–1564 (2008)
2. Ando, A., Imai, T., Kobayashi, A., Homma, S., Goto, J., Seiyama, N., Mishima, T., Kboayakawa, T., Sato, S., Onoe, K., et al.: Simultaneous Subtitling System for Broadcast News Programs with a Speech Recognizer. IEICE Transactions on Information and Systems 86(1), 15–25 (2003)
3. Moreno, P., Joerg, C., Thong, J., Glickman, O.: A Recursive Algorithm for the Forced Alignment of Very Long Audio Segments. In: Fifth International Conference on Spoken Language Processing, ISCA (1998)
4. Mnauel, J., Thong, V., Moreno, P., et al.: Speechbot: An Experimental Speech-based Search Engine for Multimedia Content on the Web. IEEE Trans. on Multi-medias 3(4), 88–96 (2002)
5. Huang, C., Hsu, W., Chang, S.: Automatic Closed Caption Alignment Based on Speech Recognition Transcripts. Technical report, Technical report, Columbia AD-VENT (2003)
6. Boulianne, G., Beaumont, J., Boisvert, M., Brousseau, J., Cardinal, P., Chapdelaine, C., Comeau, M., Ouellet, P., Osterrath, F.: Computer-Assisted Closed-Captioning of Live TV Broadcasts in French. In: Ninth International Conference on Spoken Language Processing, ISCA (2006)
7. Wheatley, B., Doddington, G., Hemphill, C., Godfrey, J., Holliman, E., McDaniel, J., Fisher, D., Inc, T., Dallas, T.: Robust Automatic Time Alignment of Orthographic Transcriptions with Unconstrained Speech. In: ICASSP 1992: IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 533–536 (1992)
8. Hosom, J.: Automatic Time Alignment of Phonemes Using Acoustic-Phonetic Information. Oregon Graduate Institute of Science and Technology (2000)
9. Rabiner, L.R., Juang, B.: Fundamentals of Speech Recognition. Pearson Education, London (2003)
10. Ney, H., Ortman, S., Aachen, T.: Progress in Dynamic Programming Search for LVCSR. Proceedings of the IEEE 88(8), 1224–1240 (2000)

11. Shao, J., Li, T., Zhang, Q., Zhao, Q., Yan, Y.: A One-Pass Real-Time Decoder Using Memory-Efficient State Network. *IEICE Transactions on Information and Systems* 91(3), 3812–3816 (2008)
12. Weintraub, M.: LVCSR Log-likelihood Ratio Scoring for Keyword Spotting. In: *ICASSP 1995: International Conference on Acoustics, Speech, and Signal Processing*, pp. 297–300 (1995)
13. Kan, M., Wang, Y., Iskandar, D., Nwe, T., Shenoy, A.: LyricAlly: Automatic Synchronization of Textual Lyrics to Acoustic Music Signals. *IEEE Transactions on Audio, Speech, and Language Processing* 16(2), 338–349 (2008)

A Perceptual Weighting Filter Based on ISP Pseudo-cepstrum and Its Application in AMR-WB

Fenglian Li and Xueying Zhang

College of Information Engineering, Taiyuan University of Technology,
Taiyuan 030024, China
gh11f1@163.com, tyzhangxy @163.com

Abstract. Wideband speech codec with bandwidth 50—7000Hz provides a significant improvement of codec speech in naturalness and intelligibility. But the spectral tilt is more pronounced in wideband speech signals due to the wide dynamic range between low and high frequencies. For solving this problem, a novel perceptual weighting filter is proposed based on the cepstral difference of Immittance Spectral Pairs (ISP) pseudo-cepstrum and linear prediction cepstral coefficients. The filter significantly compensates the spectral tilt of wideband signals that codec does not require an additional tilt compensation. The frequency response of proposed filter is consistent with the auditory masking theory. The effect of the filter to compensate the spectral tilt of wideband speech signals is much better than the perceptual weighting filter based on the cepstral difference of ISP multiplied-polynomial cepstrum and linear prediction cepstral coefficients. The effective application of the proposed filter to the adaptive multi-rate wideband (AMR-WB) speech codec indicates that the proposed filter not only efficiently compensates the spectral tilt, but also improves the objective evaluation quality values of wideband speech signals.

Keywords: ISP, ISP pseudo-cepstrum coefficients, Perceptual weighting filter, AMR-WB (Adaptive Multi-Rate Wideband) speech codec.

1 Introduction

Wideband speech codec with bandwidth 50-7000Hz provides a significant improvement in naturalness and intelligibility compared with 300-3400Hz narrowband speech codec. By lowering the low frequency cut-off from 300-50Hz, the naturalness and fullness of the speech can be improved, while extending the high frequency cut-off from 3400 to 7000Hz, the distinguishing of fricative sounds can be improved. This extended range of 50-7000Hz of wideband speech roughly corresponds to the bandwidth of speech sampled at 16kHz. Wideband speech codec can be used in video conferences, multimedia communications, and VoIP.

Algebraic Code Excited Linear Prediction (ACELP) model is used in wideband and narrowband speech codecs. It is good fit for the speech mechanism, but the coding noise should be reduced. Perceptual weighting is an important method to reduce the coding noise. It is used to shape the coding error. The basic idea behind the perceptual weighting filter is the auditory masking theory. According to the masking theory,

the coding noise in speech signal spectral peak regions would be partially or totally masked by the speech signals. However, the spectral tilt is more pronounced in wideband signals due to the wide dynamic range between low and high frequencies. The traditional perceptual weighting filter $W'(z)$ has inherent limitations in modeling the formant structure and required spectral tilt concurrently. Where filter $W'(z)$ is as follows

$$W'(z) = A(z/\gamma_1)/A(z/\gamma_2). \quad (1)$$

In (1), γ_1 、 γ_2 are perceptual weighting coefficients. $A(z)$ is a M th inverse filter, which is defined as $A(z) = 1 + \sum_{i=1}^M a_i z^{-i}$, where $\{a_i, i = 1, \dots, M\}$ are the Linear Prediction (LP) coefficients of order M .

As a wideband speech coding standard, adaptive multi-rate wideband (AMR-WB) speech codec is based on ACELP model. Its perceptual weighting filter also exploits masking properties of the human auditory system. The solution to spectral tilt problem is to introduce the pre-emphasis filter at the input, compute the LP filter $A(z)$ based on the pre-emphasised speech, and use a modified filter $W(z)$ by fixing its denominator. This structure substantially decouples the formant weighting from the tilt. The perceptual weighting filter $W(z)$ in AMR-WB [1] is as follows

$$W(z) = A(z/0.92)/(1 - 0.68z^{-1}). \quad (2)$$

But the perceptual processing is not enough for quantization noise reduction and restrain the spectral tilt phenomenon in wideband speech codec. In this paper a novel perceptual weighting filter is proposed based on Immittance Spectral Pairs (ISP) pseudo-cepstrum representation. The experiment results indicate that the proposed filter efficiently compensates the spectral tilt of wideband speech signals. The mean values of wideband-Perceptual Evaluation of Speech Quality (w-PESQ) of 9 modes in AMR-WB are improved significantly using the proposed method. Pseudo-cepstrum was introduced by H. K. Kim, K.C. Kim and H.S. Lee [2]. In [2], authors given the Linear Spectral Pairs (LSP) pseudo-cepstrum and its quefrequency-weighted version weighted pseudo-cepstrum for improving the performance of speech recognition systems. The recognition results on a set of 10 confusable syllables showed that the performance was better than that based on the LSP. In [3], authors addressed the statistical properties of the LSP pseudo-cepstrum coefficients and showed their useful application to speech recognition. In [4], an adaptive short-term postfilter based on pseudo-cepstral representation of line spectral frequencies was proposed. By applying the postfilter to several international speech coding standards, authors had reduced the complexity while obtaining a comparable performance to conventional approaches. In [5], authors moreover derived the recursive relations between LSP pseudo-cepstral representation and linear prediction polynomial cepstral coefficients.

2 The Representation of ISP Pseudo-cepstrum Coefficients

The ISP was introduced by Bistritz and Peller [6]. For a M th order LP filter, the ISPs were defined as the roots of the following polynomials (2) and (3), as well as the M th reflection coefficient k_M [7].

$$P(z) = A(z) + z^{-M} A(z^{-1}) \tag{3}$$

$$Q(z) = A(z) - z^{-M} A(z^{-1}) \tag{4}$$

In AMR-WB, the coefficients of LP filter were converted to ISP for quantization and interpolation purposes. Here M is even. Thus, (3) and (4) can be expressed as

$$P(z) = \prod_{i \in \{1,3,\dots,M-1\}} (1 - e^{j\omega_i} z^{-1})(1 - e^{-j\omega_i} z^{-1}) \tag{5}$$

$$Q(z) = (1 - z^{-2}) \prod_{i \in \{2,4,\dots,M-2\}} (1 - e^{j\omega_i} z^{-1})(1 - e^{-j\omega_i} z^{-1}) \tag{6}$$

Where $\{\omega_i\}$ is the i th Immittance Spectral Frequencies (ISF) of order M . Take the natural logarithm to (5) and (6), we can get

$$\ln Q(z) = \ln(1 - z^{-2}) + \sum_{i \in \{2,4,\dots,M-2\}} \ln(1 - e^{j\omega_i} z^{-1}) + \sum_{i \in \{2,4,\dots,M-2\}} \ln(1 - e^{-j\omega_i} z^{-1}). \tag{7}$$

$$\ln P(z) = \sum_{i \in \{1,3,\dots,M-1\}} \ln(1 - e^{j\omega_i} z^{-1}) + \sum_{i \in \{1,3,\dots,M-1\}} \ln(1 - e^{-j\omega_i} z^{-1}). \tag{8}$$

Immittance function at the glottis can be expressed as

$$I_M(z) = \frac{A(z) - z^{-M} A(z^{-1})}{A(z) + z^{-M} A(z^{-1})} = \frac{Q(z)}{P(z)}. \tag{9}$$

ISP pseudo-cepstrum coefficients $\{\hat{c}_{nl}\}$ were defined as the inverse z-transform of $\frac{1}{2} \ln[I_M(z)]$ in [8]. That is

$$\frac{1}{2} \ln[I_M(z)] = -\sum_{n=1}^{\infty} \hat{c}_{nl} z^{-n}. \tag{10}$$

At first, we take the natural logarithm to (9) and get

$$\ln(I_M(z)) = \ln Q(z) - \ln P(z). \tag{11}$$

To substitute (7) and (8) for $\ln Q(z)$ and $\ln P(z)$, then divided by 2 on both sides of (11) and take inverse z-transform, we have

$$\hat{c}_{nl} = \frac{1}{2n}(1 + (-1)^n) + \frac{1}{n} \left(\sum_{i \in \{2,4,\dots,M-2\}} \cos n\omega_i - \sum_{i \in \{1,3,\dots,M-1\}} \cos n\omega_i \right), \tag{12}$$

$n = 1, 2, \dots, M$

In (12), the first term is a constant. If we do not use the crosscorrelation of each coefficient, we can ignore the constant. Therefore, we define ISP pseudo-cepstrum coefficients $\{\hat{c}_{nl}\}$ by only using the second term on the right-hand side of (12) as

$$\hat{c}_{nl} = \frac{1}{n} \left(\sum_{i \in \{2,4,\dots,M-2\}} \cos n\omega_i - \sum_{i \in \{1,3,\dots,M-1\}} \cos n\omega_i \right), n = 1, 2, \dots, M. \tag{13}$$

From above, we know that the ISP pseudo-cepstrum $\{\hat{c}_{nl}\}$ essentially is the cepstrum of a polynomial, which is gotten through polynomial (4) divided by (3), so we call $\{\hat{c}_{nl}\}$ is an ISP divided-polynomial cepstrum. Similarly, Let polynomial (4) multiply polynomial (3), we can get the ISP multiplied-polynomial cepstrum $\{\hat{c}_{n,multiple}\}$. Here, we also ignore the constant term and get the following ISP multiplied-polynomial cepstrum

$$\hat{c}_{n,multiple} = \frac{1}{n} \sum_{i=1}^{M-1} \cos n\omega_i, n = 1, 2, \dots, M. \tag{14}$$

3 The New Perceptual Weighting Filter Based on Cepstrum Domain

Let $\{c_n\}$ expresses Linear Prediction Cepstral Coefficient (LPCC), we have

$$\ln[A(z)] = -\sum_{n=1}^{\infty} c_n z^{-n}. \tag{15}$$

Here the cepstral difference of $\{c_n\}$ and $\{\hat{c}_{nl}\}$ is defined as

$$\sum_{n=1}^{\infty} (c_n - \hat{c}_{nl}) z^{-n} = \frac{1}{2} \ln[I_M(z)] - \ln[A(z)] = \frac{1}{2} \ln \frac{Q(z)}{P(z)A^2(z)}. \tag{16}$$

To avoid (16) becomes zero as $z = \pm 1$, we move the roots of the $P(z)$, $Q(z)$ and $A(z)$ inside the unit circle. Now we can get a filter as follows

$$W_1'(z) = \frac{Q(z/\alpha_1)}{P(z/\alpha_2) A^2(z/\beta)}. \tag{17}$$

Where $0 < \alpha_1, \alpha_2, \beta < 1$. (17) has the property of perceptual weighting. It is helpful to compensate spectral tilt. When β is less than 0.5, $A^2(z/\beta)$ can be approximated as $A(z/2\beta)$ [4]. That is

$$W_1(z) = \frac{Q(z/\alpha_1)}{P(z/\alpha_2) A(z/2\beta)}. \tag{18}$$

The role of $\alpha_1, \alpha_2, \beta$ are to control the de-emphasis of formant domain. They should be determined empirically based on subjective listening results. In this paper, the optimal values are $\alpha_1 = 0.95, \alpha_2 = 0.1, \beta = 0.05$. In [8], the proposed perceptual weighting filter, which was based on the cepstral difference of $\{c_n\}$ and $\{\hat{c}_{n,multiple}\}$, had the following form

$$W_2(z) = \frac{Q(z/\alpha_1') P(z/\alpha_2')}{A(z/2\beta')}. \tag{19}$$

Where $\alpha_1' = 0.95, \alpha_2' = 0.15, \beta' = 0.35$.

Fig. 1 gives the 16th-order LPC spectral envelopes of wideband speech signals, the frequency responses of $W(z)$, $W_1(z)$ and $W_2(z)$, respectively. Where, the above thick solid curves is the original speech signals LPC spectral envelopes. The thin solid curves is the frequency responses of $W(z)$. The dash curves is the frequency responses of $W_1(z)$. The dash-dot curves is the frequency responses of $W_2(z)$. We can find that the spectral tilt phenomenon of wideband speech signals is obvious. $W_1(z)$ and $W_2(z)$ have spectral valleys in speech signal formant domains and have spectral formants in speech signal valley domains. It is consistent with the masking effect of human hearing system. Compared with $W_2(z)$, the frequency response of $W_1(z)$ has bigger spectral formants. The spectral formants of high frequency domain are improved moreover. The effect to compensate spectral tilt is more enough.

Fig. 2 shows the magnitude spectra and LPC spectral envelopes of original, $W_1(z)$ and $W_2(z)$ weighted speech signals, respectively. The above three curves are the magnitude spectra. The under three curves are the 16th-order LPC spectral envelopes. In Fig.2, two thick solid curves express original speech signals magnitude spectra and its LPC spectral envelope, respectively. Two thin solid curves express $W_1(z)$ weighted speech signals magnitude spectra and its LPC spectral envelope, respectively. Two x curves express $W_2(z)$ weighted speech signals

magnitude spectra and its LPC spectral envelope, respectively. It is obvious that the proposed filter $W_1(z)$ can more effectively compensate spectral tilt of wideband speech signals than $W_2(z)$ do.

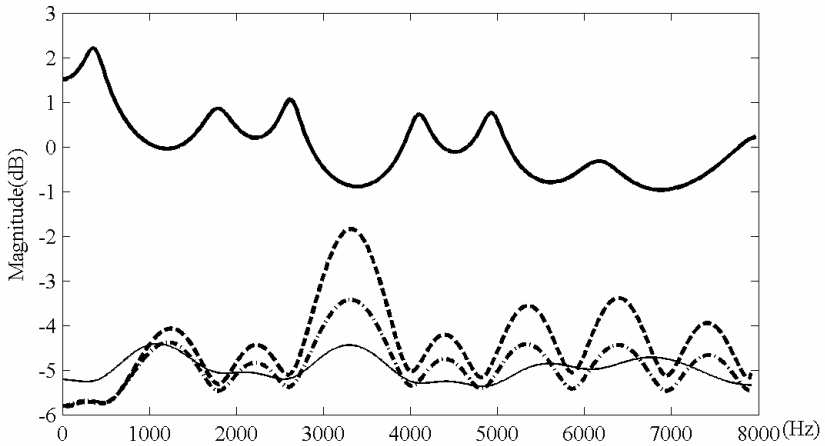


Fig. 1. Comparison of original speech signals LPC spectral envelope (*thick solid curves*) and the frequency responses of $W(z)$ (*thin solid curves*), $W_1(z)$ (*dash curves*) and $W_2(z)$ (*dash-dot curves*)

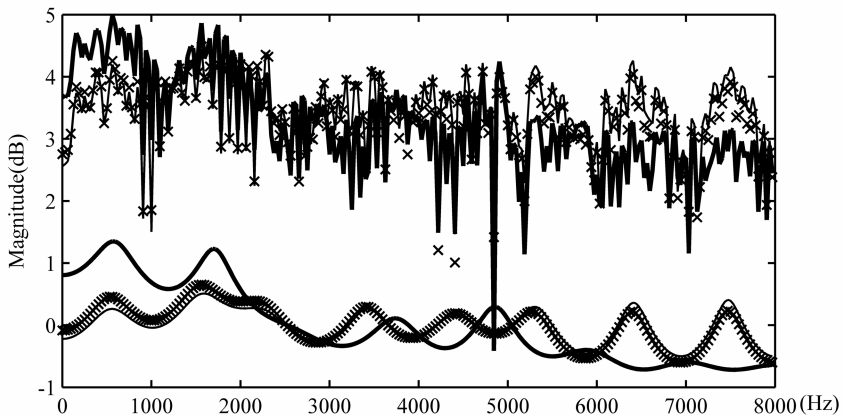


Fig. 2. Comparison of magnitude spectra (above three curves) and LPC spectral envelopes (under three curves) of original speech signals (*thick solid curves*), $W_1(z)$ (*thin solid curves*) and $W_2(z)$ (*x marked curves*)

4 Simulations

Simulations are based on 20 male speech sentences and 20 female speech sentences from the TIMIT database. The objective scores are evaluated using the w-PESQ (wide-band Perceptual Evaluation of Speech Quality) from ITU-T P.862.2. Simulations are carried out with the ITU-T G.722.2 AMR-WB speech codec. Table 1 and table 2 respectively give male and female speech sentences average w-PESQ of 9 modes in AMR-WB. They include w-PESQ of unweighted decoding speech signals, $W(z)$ and $W_1(z)$ weighted decoding speech signals. Table 3 and table 4 respectively give male and female speech sentences average w-PESQ improvements of $W(z)$ and $W_1(z)$ weighted decoding speech signals. Compared to unweighted signals, the results in Table 1 to table 4 show that $W(z)$ and $W_1(z)$ weighted signals w-PESQ greatly increase.

To male sentences, the unweighted decoding speech signals average w-PESQ of 9 modes is only 3.439, whereas $W(z)$ and $W_1(z)$ weighted signals average w-PESQ increase to 4.153 and 4.187, respectively. The improvements are 0.714 and 0.748, respectively. To female sentences, the unweighted decoding speech signals average w-PESQ of 9 modes is only 3.143, whereas $W(z)$ and $W_1(z)$ weighted signals average w-PESQ increase to 4.047 and 4.085, respectively. The improvements are 0.904 and 0.942, respectively. Compared to $W(z)$ weighted signals, $W_1(z)$ weighted signals average w-PESQ improvements of male and female sentences are 0.034 and 0.038, respectively. These indicate that the proposed perceptual weighting filter $W_1(z)$ is good fit for perceptual weighting in AMR-WB.

Table 1. The mean w-PESQ scores of unweighted, $W(z)$ and $W_1(z)$ weighted 20 male speech sentences

Mode	w-PESQ		
	Unweighted Speech	$W(z)$ weighted	$W_1(z)$ weighted
0	2.841	3.744	3.794
1	3.096	3.967	4.012
2	3.382	4.165	4.189
3	3.476	4.192	4.227
4	3.509	4.221	4.265
5	3.601	4.269	4.282
6	3.656	4.257	4.294
7	3.705	4.284	4.319
8	3.693	4.278	4.304
Average	3.439	4.153	4.187

Table 2. The mean w-PESQ scores of unweighted, $W(z)$ and $W_1(z)$ weighted 20 female speech sentences

Mode	w-PESQ		
	Unweighted Speech	$W(z)$ weighted	$W_1(z)$ weighted
0	2.528	3.557	3.616
1	2.776	3.803	3.857
2	3.071	4.065	4.100
3	3.138	4.107	4.139
4	3.195	4.124	4.164
5	3.313	4.168	4.198
6	3.352	4.177	4.201
7	3.469	4.209	4.246
8	3.441	4.218	4.241
Average	3.143	4.047	4.085

Table 3. $W(z)$ and $W_1(z)$ weighted 20 male speech sentences mean w-PESQ improvements

Mode	Improved w-PESQ		
	$W(z)$ weighted	$W_1(z)$ weighted	$W_1(z)$ exceeding $W(z)$
0	0.903	0.953	0.050
1	0.871	0.916	0.045
2	0.783	0.807	0.024
3	0.716	0.751	0.035
4	0.712	0.756	0.044
5	0.668	0.681	0.013
6	0.601	0.638	0.037
7	0.579	0.614	0.035
8	0.585	0.611	0.026
Average	0.714	0.748	0.034

Table 4. $W(z)$ and $W_1(z)$ weighted 20 female speech sentences mean w-PESQ improvements

Mode	Improved w-PESQ		
	$W(z)$ weighted	$W_1(z)$ weighted	$W_1(z)$ exceeding $W(z)$
0	1.029	1.088	0.059
1	1.027	1.081	0.054
2	0.994	1.029	0.035
3	0.969	1.001	0.032
4	0.929	0.969	0.040
5	0.855	0.885	0.030
6	0.825	0.849	0.024
7	0.740	0.777	0.037
8	0.777	0.800	0.023
Average	0.904	0.942	0.038

5 Conclusions

In this paper, the concept of ISP pseudo-cepstrum and the equation of ISP pseudo-cepstrum coefficients are introduced at first. Then a new perceptual weighting filter is designed based on ISP pseudo-cepstrum domain. Simulation results indicate that the designed filter is helpful to compensate the spectral tilt and improve wideband speech signal perceptual weighting quality. The objective assessment scores obtained using w-PESQ can be improved significantly by proposed method.

Acknowledgments. The work was supported by Shanxi Province Natural Science Foundation of China under the grant No.20051039.

References

1. ITU-T Recommendation G.722.2.: Wideband Coding of Speech at Around 16kbit/s Using Adaptive Multi-rate Wideband (2003)
2. Kim, H.K., Kim, K.C., Lee, H.S.: Enhanced Distance Measure for LSP-based Speech Recognition. Electronics Letters 29(16), 1463–1465 (1993)

3. Kim, H.K., Choi, S.H., Lee, H.S.: On Approximating Line Spectral Frequencies to LPC Cepstral Coefficients. *IEEE Transactions on Speech and Audio Processing* 8(2), 195–199 (2000)
4. Kim, H.K., Kang, H.G.: An Adaptive Short-term Postfilter based on Pseudo-cepstral Representation of Line Spectral Frequencies. *J. Speech Communication* 37, 335–348 (2002)
5. Kim, H.K., Choi, S.H.: Cepstral Domain Interpretations of Line Spectral Frequencies. *J. Signal Processing* 88, 756–760 (2008)
6. Bistriz, Y., Peller, S.: Immittance Spectral Pairs (ISP) for Speech Encoding. In: *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, pp. II-9–II-12 (1993)
7. Stephen, S., Kuldip, K., Paliwal: A Comparative Study of LPC Parameter Representations and Quantization Scheme for Wideband Speech Coding. *J. Digital Signal Processing* 17, 114–137 (2007)
8. Li, F.L., Zhang, X.Y.: A Perceptual Weighting Filter on Cepstrum Domain for Wideband Speech Codecs. In: *2008 11th IEEE International Conference on Communication Technology Proceedings*, pp. 688–691. IEEE Press, Hangzhou (2008)

Video Fingerprinting by Using Boosted Features

Huicheng Lian¹ and Jing Xu²

¹ School of Computer Engineering and Science, Shanghai University,
P. Box 147, No. 149 Yanchang Road, Shanghai, 200072, China

² Shanghai Jiao Tong University Library, Shanghai, 200240, China
lianhc@shu.edu.cn, xujing@lib.sjtu.edu.cn

Abstract. In this paper, we present a novel approach for video fingerprinting by using boosted Harr-like features and direct hashing. Through employing a pairwise boosting method on a large set of features, our system can learn the top- M discriminative filters that are enable to efficiently extracting video fingerprints. During query phase, we retrieve video clips by using a fast and accurate direct hashing, which minimizes perceptual Hamming distance between queries and a large database of pre-computed fingerprints. To demonstrate the superiority of our method, we also implement four other fingerprinting methods for comparisons. The experimental results indicate that our proposed method can significantly outperform those four methods in video retrieval.

1 Introduction

Fingerprints are defined as perceptual features or short summaries of a multimedia object, and the goal of fingerprinting is to provide fast and reliable methods for content identification [1][2]. Specifically, video fingerprints are feature vectors that uniquely characterize one video clip from another, and the goal of video fingerprinting is to identify a given video query in a database by measuring the distance between the query fingerprint and the fingerprints in the database [2]. Promising applications of video fingerprinting are filtering file-sharing services, broadcast monitoring, automated indexing of large-scale video archives, etc [2].

In these recently years, many video fingerprinting methods have been proposed. For example, Oosteevn *et al* proposed a differential block luminance algorithm for video fingerprints extraction [1]. They also introduced a structure for very efficient searching in a large fingerprinting database, which belongs to perceptual hash methods [1]. Lee and Yoo proposed a video fingerprinting method based on the centroid of gradient orientations [2][3]. The centroid of gradient orientations was used due to its pairwise independence and robustness against common video processing steps [2][3]. Kim *et al* calculated a binary image signature for each key frame by averaging Y component in YCbCr color layout for each macro blocks. Extreme quantization then was applied to obtain a binary image signature of a given frame [4]. Ramachandra *et al* proposed a 3D video fingerprinting by using scale invariant feature descriptor (SIFT) descriptors [5].

Sarkar *et al* presented a Compact Fourier-Mellin Transform (CFMT) for fingerprints extraction and compared it with SIFT and YCbCr histogram-based feature methods [6].

Another technique called audio fingerprinting was proposed for audio's retrieval early. In 2002, Oosteevn *et al* proposed an approach to convert audio signal into 2D time-frequency image (spectrogram), using the short-term Fourier transform (STFT) for audio fingerprints extraction [7]. Based on this method, Ke *et al* proposed to employ a pairwise boosting on a large set of Viola-Jones features or called Harr-like features [10], to learn compact, discriminative, local descriptors for spectrogram [8]. Ke's computer vision techniques for audio fingerprint made a significant contribution to audio retrieval. Following their line, Kim and Yoo proposed a boosted audio fingerprint method based on spectral subband moments [9]. Baluja and Covella proposed a audio fingerprinting called waveprint [11] [12]. A difference between Ke's method and Baluja's method is that, Ke used a boosting method to choose top- M filters while Baluja extracted the top- M Haar-wavelets according to their magnitudes in a spectrogram.

Motivated by thinking of Ke [8], we propose a modified pairwise boosting to learn top- M discriminative filters for video fingerprinting. Specifically, a large set of Viola-Jones features [10] are generated from frames of matching and non-matching video clips firstly, then a pairwise boosting is employed to choose the top- M features by iteratively constructing and finding weak classifiers with minimal error. After obtaining the top- M filters, we can process all videos by using these filters in the same way. During retrieval, the query fingerprints are input into a direct hashing retrieval system for fast searching.

2 Preprocessing

To reduce the noises produced from video encoding (such as bit rate, format) or editing (such as resize, frame rate), a preprocessing is much necessary. We preprocess all videos as following steps: (1) resize them to $W \times L$ pixels, for example $W = 160$ and $L = 120$ in this paper. (2) change the frame rate to a fixed frame rate, for example, 6 frames per second in this paper. (3) bit rate of encoding is fixed to a solid value. After these procedures, videos are conducted into video fingerprinting encoding module, which will be detailed in next sections.

3 Pairwise Boosting for Filter Selection

In 2001, Viola and Jones introduced Haar wavelet-like features for face detection [10]. The simple features used are differences among convolutions of Haar basis functions on rectangular regions of images. Within any image sub-window the total number of Harr-like features is very large. In order to ensure fast classification, a small set of critical features should be focused on. A feature selection was achieved through a AdaBoost procedure. The weak learner was constrained so that each weak classifier returned can depend on only a single feature [10]. As a

result, each stage of the boosting process, which selects a new weak classifier, can be viewed as a feature selection processing.

Ke *et al* [8] proposed to employ a pairwise boosting on a large set of Viola-Jones features [10], to learn compact, discriminative, local descriptors on spectrograms [8]. The top- M learned features, which are distinctive while resistant to expected distortions, were chosen to be filters for audio fingerprinting extraction. Similarly, there is a very large set of Harr-like features within every video frame. As we can imagine, different features (as filters) have different distinctive abilities. Filters with large width and length can more robust to certain video distortions, while filters with short width and length can capture discriminative information that the former filters cannot. The learned filter family should be

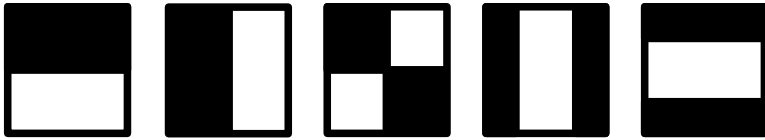


Fig. 1. Candidate Harr-like features

– **Pairwise Boosting**

– **Input:** sequence of n samples $\langle(x_{11}, x_{12})\rangle, \langle(x_{21}, x_{22})\rangle, \dots, \langle(x_{n1}, x_{n2})\rangle$, each with label $y_i \in \{-1, 1\}$

– **Initialize** $w_i = \frac{1}{n}$, $i = 1, \dots, n$

– **For** $m=1, \dots, M$

1 find the hypothesis $h_m(x_1, x_2)$ that minimizes weighted error over distribution w , where

$$h_m(x_1, x_2) = \text{sgn}[(f_m(x_1) - t_m)(f_m(x_2) - t_m)]$$

for filter f_m and threshold t_m

2 calculate weighted error:

$$\text{err}_m = \sum_{i=1}^n w_i \cdot \delta(h_m(x_{i1}, x_{i2}) \neq y_i)$$

3 assign confidence to h_m :

$$c_m = \frac{1}{2} \cdot \log\left(\frac{1 - \text{err}_m}{\text{err}_m}\right)$$

4 update weights for matching pairs:

$$w_i = w_i \times \begin{cases} e^{c_m} & \text{if } h_m(x_{i1}, x_{i2}) \neq y_i \\ e^{-c_m} & \text{if } h_m(x_{i1}, x_{i2}) = y_i \end{cases}$$

5 normalize weights such that

$$\sum_{i:y_i=-1}^n w_i = \sum_{i:y_i=1}^n w_i = \frac{1}{2}$$

– **Finally hypothesis:**

$$H(x_1, x_2) = \text{sgn}\left(\sum_{m=1}^M c_m \cdot h_m(x_1, x_2)\right)$$

Fig. 2. Pairwise Boosting Algorithm

able to capture characteristics that can be distinctive. This is what we want to do by using a pairwise boosting algorithm.

One way to learn a small set of critical filters from a large set of candidate filters is pairwise boosting method described in [8]. A small-modified version is proposed in our method, which is shown in Figure 2. Notice that, as analyzed by Ke *et al* [8], non-matching examples would be incorrectly classified as matching at least half of the time for a sufficient large sample size. So, the pairwise boosting method is actually an asymmetric algorithm, in which only the matching pairs ($y=1$) are re-weighted. The weights of matching pairs and non-matching pairs are normalized such that the sum of each is equal to one-half [8].

Five Harr-like features employed are shown in Fig. 1. The width and height of a frame were divided into 16 and 12 units, respectively, with every 10 pixels being one unit. Each filter type can vary in x -axis from 1 to 15 and in y -axis from 1 to 11. The width varies from 1 to 16 and height varies from 1 to 12. This results in a set of 53,040 candidate filters for selection. We random selected 10000 pairs from 60 videos for pairwise boosting. The procure took us 18 hours to finish the iterations. The top $M = 32$ discriminative filters were selected for fingerprinting extraction. The first four Haar-like features selected are shown in Figure 3. From this figure, we can see that features with large width and length are firstly selected. They are considered to be more robust to certain distortions.

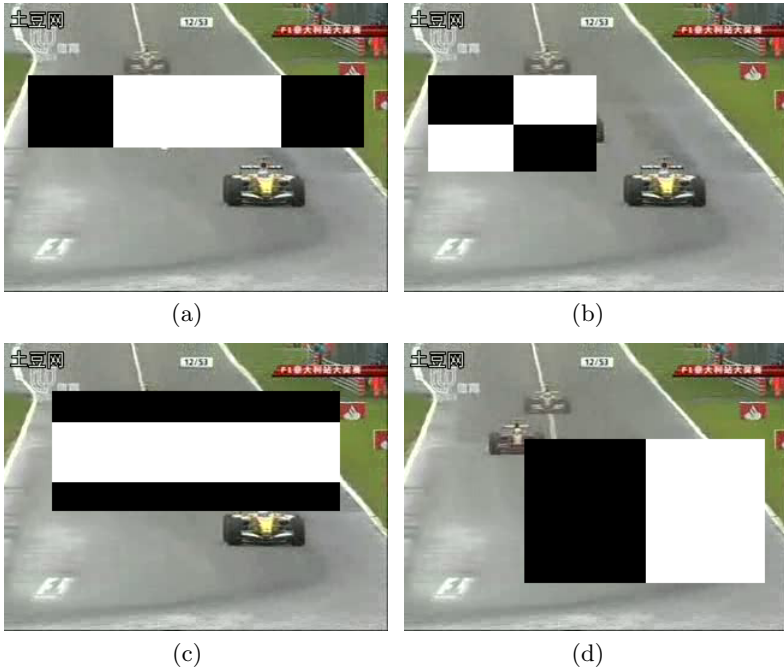


Fig. 3. The top four Harr features selected

4 Retrieval

Fast retrieval is another important task for fingerprinting system. In Oostveen’s papers [1][7], they proposed a structure for very efficient searching in a large fingerprint database. The fingerprint database contains a lookup table (LUT) with all possible 32 bit sub-fingerprints as an entry. Every entry points to a list with pointers to the positions in the real fingerprint lists. In practical systems, with limited memory, a lookup table containing 2^{32} entries is often not feasible. Therefore, in practice, a hash table is used instead of a lookup table. Ke *et al* called this direct hashing method [8].

Another hash method named Local Sensitive Hashing (LSH) proposed in [14]. However, in Ke’s paper [8], he found that his audio descriptors are so robust that direct hashing, using a classical hash table, greatly reduced running time without significantly impacting accuracy. Actually, as described in [14][11], LSH performs a series of hashes, each of which examines only a portion of the sub-fingerprints, so it must be more slow and space-consumed than direct hashing. In our system, we employed direct hashing as our retrieval method.

The direct hashing [8] can be summarized as following three steps: (1) Matching all the nearly similar neighbors of every sub-fingerprints in one query, from the whole LUT pre-built, (2) filtering those matched sub-fingerprints who do not have a same file ID of most, (3) verifying all filtered query by using a Hamming distance metric [8], where a smaller distance means a more perceptual similarity.

5 Evaluation

To obtain an evaluation, we choose following methods for comparison: Oostveen’s method [1], Kim’ method [4], Lee and Yoo’s method [2], manual Harr-like features method and Harr-like features’ boosting method proposed in this paper. Experimental setup and performance analysis are detailed in next sub-sections.

5.1 Experimental Setup

To Oostveen’s method [1], we use $R = 4, C = 9$ and $\alpha = 0$. The reason why we do not use $\alpha = 0.95$ will be analyzed in the later. To Kim’s method [4], we employ a 4 columns and 8 rows dividing method to make sure a sub-fingerprint we obtain is 32 bits. To Lee and Yoo’s method [2], we also employ a 4 columns and 8 rows dividing method. A little difference is that we normalize centroid of gradient orientation to 0 or 1 but not $[-\pi/2, \pi/2)$. The reason is that, we can not obtain 32 bits for a frame if using continuous values. To manual Harr-like features method, we select six proportional spacing blocks and perform five Harr-like features on them respectively. Then we perform the first two Harr-like features on the whole frame. Totally, we obtain 32 bits for one frame. Other than these methods, actually, we also consider many methods, such as color histogram extraction methods, optical flow extraction methods and SIFT methods. However, their performances are too bad that we do not show them in our experimental results.

We downloaded 245 video files (FLV format) from Internet, which totally occupied 5.36 Gigabytes, for experiments. In these video files, there are 11 video groups downloaded from 11 different content channels from Tudou.com¹. Each group contains 10 videos, and these videos are perceptually same to each other in one group. However, they are various in many aspects, such as, bit rate, caption, logo, resolution and other distortions or noises, generated when videos are transferred from one to another. The other 135 videos were validated manually to be different from any videos in the 11 groups. We randomly selected one videos from every group. The total 11 videos from the 11 groups were used as a referred database for retrieval. The rest 99 videos from 11 groups were used as a positive query set, and the other 135 videos were used as a negative query set. Totally, 15,418 small clips were generated from positive and negative query sets for query, with each clip containing $6 \times 60 = 360$ sub-fingerprints. The five video fingerprinting methods are compared and shown in next subsection, with the same experimental setups described above.

5.2 Performance Analysis

The performances of the fingerprinting methods are plotted using a receiver operating characteristics (ROC) curve. This is a plot of the false negative rate (FNR) versus the false positive rate (FPR). Let N_p be the total number of positive clips, and N_n the total number of negative clips. With FN the number of false negatives and FP the number of false positives, we have FNR and FPR as follows:

$$FNR = \frac{FN}{N_p} \times 100\%, \text{ and } FPR = \frac{FP}{N_n} \times 100\%$$

We plot points representing these rates on a two dimensional graph with changing threshold value from minimum value to maximum value. The curves of five fingerprinting methods are shown in Figure 4.

From this figure, we can see that, Lee and Yoo's method [3] performs worst among these methods. It is possible caused by our discrete normalization on it. But if not, continuous centroid of gradient orientation values will cause a saving space 32 times to discrete normalization. A more serious thing is, the fast direct hashing retrieval would not works on such continuous values, since it computes Hamming distance on bits but not values. Manual Harr-like features method has a little better performance than Lee and Yoo's method. However it does not be a very good performance in this experiment. This means that we can select Harr-like features by other methods, such as boosting, but not only by a manual manner. Oosten's method [1] performs better than the former two methods in our experiment. Notice that here α is set to 0 but not 0.95. Actually, we also try $\alpha = 0.95$, but find it is very bad in the experiment. The reason we considered it is, if using 0.95, the threshold will not be zero statistically when calculating fingerprinting bits, however in [1] 0.95 is used. This will bring a bias

¹ <http://www.tudou.com>

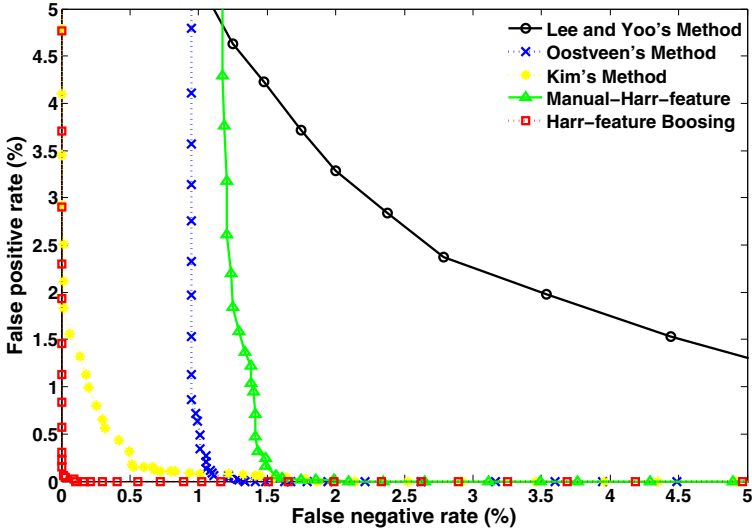


Fig. 4. ROC of four fingerprinting methods

for calculating fingerprinting bits and therefore result in a bad discrimination. Kim’s method [4] looks very simply but it is efficient in our experiment. The average Y component in YCbCr color layout of the divided blocks can roughly catch distributing information in each frame. And when one video clip is represented as a "group" of frame, time-domain information among frames actually be included inside sub-fingerprints.

From Figure 4, we can see that, our proposed method performs a best performance among all five methods. For a more clear display, we output the *FNR* and *FPR* pairs who have minimum distances between themselves. The *FNR* of former four methods are 2.78%, 1.34%, 0.94%, 0.42%, when *FPR* values of them are 2.37%, 1.37%, 0.86%, and 0.43%. While the *FNR* and *FPR* of our proposed method are only 0.03% and 0.04%, respectively. This demonstrates that our proposed method take a significant improvement to others methods. We consider the reasons as follows: (1) Harr-like features can capture video object’s features, in a form of calculating various differential values in frame blocks, (2) Filters with large width and height are more robust to certain distortions, but filters with short width and height can capture discriminative information that the former cannot, (3) Pairwise boosting helps to select *M* filters who have most distinctive abilities, or whose combination can have a most distinctive ability.

Figure 5 presents the retrieval rates of proposed methods with various querying lengths of 10, 20, 40, and 60 seconds, corresponding to 60, 120, 240, and 360 sub-fingerprints, respectively. From this figure we can see that, the longer query largely improves the retrieval results. However, users usually wish to achieve the desired accuracy using as short as a query as possible. So, there is trade-off we

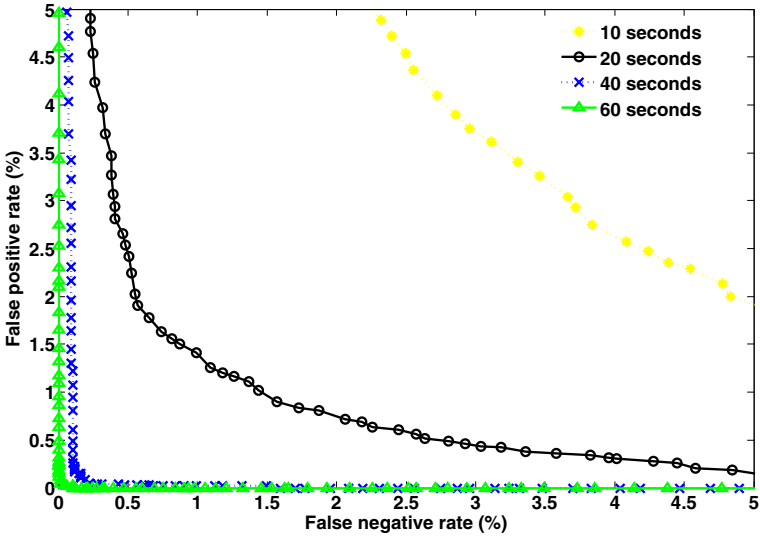


Fig. 5. ROC of different lengths

need to decide between query length and accuracy. In practical applications, we proposed using 360 sub-fingerprints for retrieval.

6 Conclusions

We have proposed a fast and accurate fingerprinting approach for video identification. Firstly, a preprocessing was proposed to reduce video noises that may introduced from various transformations. This preprocessing includes scaling to an uniform size, fixing to a solid frame rate and fixing to a solid bit rate. Then, a pairwise boosting method was employed to finish feature selecting for fingerprinting on the preprocessed videos. The small set of filters (i.e. features), selected by boosting, were demonstrated to be efficient at capturing video frame's object information in a very compact way. As a result, a video frame can be presented only by a 32 bits number. This compact and discriminative representing way allows a very fast retrieval for querying of video clips. From experimental results, we can see that the proposed method significantly outperforms Oostveen's method, Kim' method, Lee and Yoo's method, and a manual Harr-like features method. In future work, we plan to study a larger scale indexing method and more accurate extraction methods for our video retrieval.

Acknowledgment

This work is supported by Shanghai Leading Academic Discipline Project, Project Number: J50103, and also supported by the Excellent Young Teacher Foundation of Shanghai via the grant shu08068.

References

1. Oostveen, J., Kalker, T., Haitsma, J.: Feature Extraction and a Database Strategy for Video Fingerprinting. In: Chang, S.-K., Chen, Z., Lee, S.-Y. (eds.) VISUAL 2002. LNCS, vol. 2314, pp. 67–81. Springer, Heidelberg (2002)
2. Lee, S., Yoo, C.D.: Robust Video Fingerprinting for Content-Based Video Identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, 707–711 (2002)
3. Lee, S., Yoo, C.D.: Video Fingerprinting Based on Centroids of Gradient Orientations. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2006)*, vol. 2, pp. 14–19 (2006)
4. Kim, H.S., Lee, J., Liu, H.B., Lee, D.W.: Video Linkage: Group Based Copied Video Detection. In: *ACM Int'l Conf. on Image and Video Retrieval (CIVR)*, Niagara Falls, Canada, July 2008, pp. 397–406 (2008)
5. Ramachandra, V., Zwicker, M., Nguyen, T.: 3D Video Fingerprinting. In: *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video 2008*, pp. 28–30, 81–84 (2008)
6. Sarkar, A., Ghosh, P., Moxley, E., Manjunath, B.S.: Video Fingerprinting: Features for Duplicate and Similar Video Detection and Query-based Video Retrieval. In: *Proc. SPIE - Multimedia Content Access: Algorithms and Systems II*, San Jose, California (January 2008)
7. Oostveen, J., Kalker, T., Haitsma, J.: An efficient Database Search Strategy for Audio Fingerprinting. In: *IEEE Workshop on Multimedia Signal Processing 2002*, December 9–11, pp. 178–181 (2002)
8. Ke, Y., Hoiem, D., Sukthankar, R.: Computer Vision for Music Identification. In: *Proceedings of Computer Vision and Pattern Recognition 2005*, vol. 2, pp. 1184–1192 (2005)
9. Kim, S., Yoo, C.D.: Boosted Binary Audio Fingerprint Based on Spectral Subband Moments. In: *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2007*, April 15–20, pp. 241–244 (2007)
10. Viola, P., Jones, M.: Rapid Object Detection Using a Boosted Cascade of Simple Features. In: *Proceedings of Computer Vision and Pattern Recognition*, vol. 1, pp. 511–518 (2001)
11. Baluja, S., Covell, M.: Audio Fingerprinting: Combining Computer Vision & Data Stream Processing. In: *IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2007*, April 15–20, pp. 213–216 (2007)
12. Baluja, S., Covell, M.: Waveprint: Efficient wavelet-based audio fingerprinting. *Pattern Recognition* 41(11), 3467–3480 (2008)
13. Schapire, R.E.: A Brief Introduction to Boosting. In: *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence 1999*, San Francisco, CA, USA, pp. 1401–1406 (1999)
14. Indyk, P., Motwani, R.: Approximate Nearest Neighbor- towards Removing the Curse of Dimensionality. In: *Proceedings of Symposium on Theory of Computing*, pp. 604–613 (1998)

Reference Signal Impact on EEG Energy

Sanqing Hu^{1,2}, Matt Stead¹, Hualou Liang², and Gregory A. Worrell¹

¹ Department of Neurology, Division of Epilepsy and Electroencephalography
Mayo Clinic, Rochester, MN 55905, USA

² School of Biomedical Engineering, Drexel University, Philadelphia, PA 19104, USA
worrell.gregory@mayo.edu

Abstract. A reference is required to record electroencephalography (EEG) signals, and therefore the reference signal can effect any quantitative EEG analysis. In this study, we investigate the impact of reference signal amplitude on a commonly used quantitative measure of the EEG, the signal energy. We show that: (i) when the reference signal and the non-referential signal have negative correlation, the energy of the referential signal will monotonically increase as the amplitude of the reference signal increases from 0 to ∞ . (ii) When the reference signal and the non-referential signal have positive correlation, energy of the referential signal first decreases to some nonnegative value and then increases as the amplitude of the reference signal increases from 0 to ∞ . In general, the reference signal may decrease or increase energy values. But a reference signal with higher relative amplitude will surely increase energy values. In [1], we developed a method to identify and extract the reference signal contribution to EEG recordings. Here we apply this approach to referential EEG recorded from human subjects and directly investigate the contribution of recording reference on energy and show that the reference signal may have a significant effect on energy values.

Keywords: Scalp reference signal, Referential EEG, Corrected EEG, Bipolar EEG, Energy.

1 Introduction

It is estimated that 50 million people world-wide have epilepsy. For patients the unpredictability of seizure occurrence remains one of the most devastating aspects of epilepsy. The possibility of predicting the onset of seizures is clinically very attractive. If prediction were possible patients could be given a warning of impending seizures and take evasive actions to limit the chance of injury or therapy could be given. During the last decade numerous methods have been proposed to predict epileptic seizures, extracting quantitative features from EEG recordings, based concepts including nonlinear dynamics and chaos, signal energy, phase synchronization, and wavelet transform, etc (see [2] and the references therein). A major advantage of energy-based measures [3]–[6] is that they are computationally efficient, easy to relate to raw data, and are easily implemented in implantable devices [3]. However, unfortunately, the results to date using quantitative EEG are not sufficient for clinical usable devices. One possible confound is EEG signal itself. In fact, since the EEG reflects the difference between electrical potentials measured at two different electrodes the signals are always confounded by the contribution

from two recording locations, the electrode of interest and the reference electrode. The vast majority of clinical and research EEG recordings, both scalp and intracranial, are obtained using a common cephalic reference. As a result, the reference signal generally has an effect on the EEG.

In the literature, almost all EEG analysis are based on either common referential EEG recordings or common reference-free EEG recordings such as bipolar EEG, average common reference EEG and Laplacian EEG. Recently the potential pitfalls of using common referential EEG recordings for correlation, coherence and phase synchrony analysis have been established [7]. The potential pitfalls associated with the use of bipolar EEG for coherence analysis are also recognized [8]. Although the average reference EEG and Laplacian EEG are reference-free, problems with their use for synchronization analysis is also recognized [7].

In our recent work [1] we proposed two methods to extract the scalp reference signal from clinical multi-channel iEEG recordings based on independent component analysis [9], and stated why the obtained signal is a “good” estimation of the real reference signal. The corrected EEG, or true reference-free EEG, can now be obtained by removing the reference signal. In this study, we focus on reference effect on EEG energy.

2 Methods and Material

Given one time-series $x(t)$, energy of $x(t)$ is defined as

$$E_x = E[x^2], \quad (1)$$

that is, the variance of x where $E[\cdot]$ is the expected value of one random variable.

Now we investigate the effect of recording reference on energy of EEG signal. Let $R(t) = Ar(t)$ denote the potential signal at the reference electrode where coefficient $A > 0$, and b be the potential signal at the intracranial or scalp electrode. Now let $x(t)$ denote $b(t)$ referenced to $R(t)$, that is, $x(t) = R(t) - b(t) = Ar(t) - b(t)$. Now we aim to discuss the effect of $R(t)$ on energy of non-referential signal $b(t)$ as measured from referential signal $x(t)$.

Putting $x(t) = R(t) - b(t)$ into (1), we obtain

$$\begin{aligned} E_x &= E[x^2] = E[(R - b)^2] = E[(Ar - b)^2] = E[r^2]A^2 - 2E[rb]A + E[b^2] \\ &\triangleq E_x(A). \end{aligned} \quad (2)$$

Thus, $E_x(A)$ is a function of coefficient A where $A > 0$. When $E[rb] \leq 0$ which means r and b have negative correlation, it is easy to see that $E_x(A)$ is a monotonically increasing function as A varies in $[0, +\infty)$. So, the reference signal always increases energy value in this case. Figure 1(a) shows examples of the function $E_x(A)$ according to negative correlations of r and b . When $E[rb] > 0$ which means r and b have positive correlation, based on (2) one can get that $E_x(A)$ has the minimum value of

$$\frac{E[r^2] \times E[b^2] - (E[rb])^2}{E[r^2]} \geq 0$$

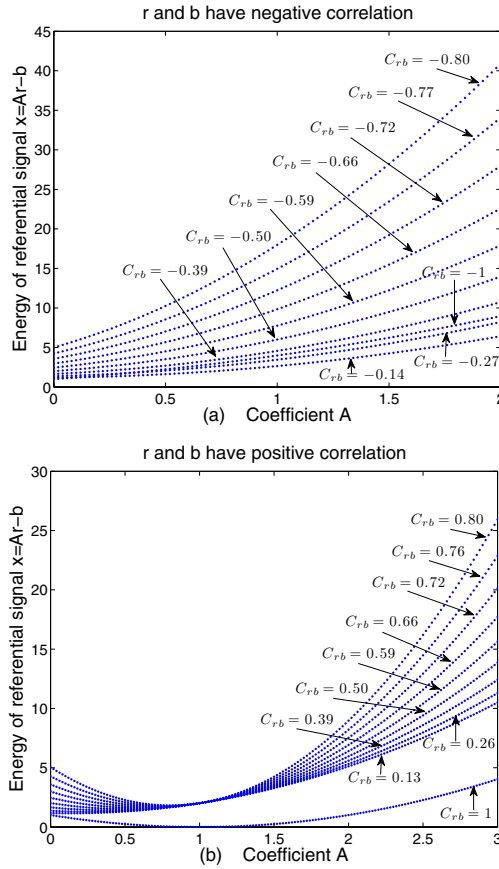


Fig. 1. (a) Energy of referential signal as a function of coefficient A where the reference signal and non-referential signals have negative correlation. (b) Energy of referential signal as a function of coefficient A where the reference signal and non-referential signals have positive correlation. In (a) and (b) r and b are randomly generated with some correlation coefficient.

at the critical point $A^* = E[rb]/E[r^2]$. So, $E_x(A)$ is a monotonically decreasing function in $(0, A^*]$ and then starts to monotonically increase in $(A^*, +\infty)$. Figure 1-(b) shows examples of the function $E_x(A)$ according to positive correlations of r and b . From Figure 1-(b) one can see that the reference signal with high relative amplitude will always increase energy value of non-referential signal.

Intracranial EEGs and scalp EEGs were recorded from one patient being monitored for epilepsy surgery by using a stainless steel suture placed in the vertex region of the scalp, midline between the Cz and Fz electrode positions (international 10–20) as a common reference. The data were acquired on an XLTekTM EEG 128 system that digitizes each channel at 500 Hz using a pre-digitization analog high-pass filter at 0.01 Hz and low pass filter at 125 Hz. We calculate energy for every sliding window of 1 second without overlap.

3 Results

The patient underwent iEEG monitoring using 16 contact depth electrodes (LTD1~LTD8 and RTD1~RTD8) placed within the right and left medial temporal lobes and 20 surface electrodes at F7, T7, P7, Fp1, F3, C3, P3, O1, Fpz, Fz, Cz, Pz, Oz, Fp2, F4, C4, P4, F8, T8, P8 (international 10–20) recorded from the same vertex reference electrode all sampled at 500Hz. Each data segment analyzed contains 50000 samples (100 seconds) and was obtained in the quiet awake resting state. See patient 2 in [1] for detail. The reference signal (R2) in Figure 2(a) was calculated based on the second method in [1] by using the entire time period (100 seconds) and all 16 iEEG channels. In [1] we explained why R2 is a “good” estimation of the real reference. In Figure 2(a), iEEG (RTD5 and RTD6) and scalp EEG (Cz and Pz) are plotted where only 10 seconds of 100 seconds are shown representatively. cRTD5, cRTD6, cCz and cPz are corrected EEGs obtained by subtracting R2 from the corresponding raw EEGs. It can be observed that raw RTD5 and RTD6 were contaminated by muscle artifacts that are removed in the corrected iEEG. This shows that i) muscle artifacts in the referential iEEG come from the reference signal in this case, and ii) the reference signal was mostly removed using the approach described in [1]. Bipolar montage iEEG (RTD5–RTD6) is also plotted in Figure 2(a) and is all muscle artifact-free. This further verifies that in this example the artifacts are from the common reference signal, and therefore removed in the bipolar montage. It is notable that (i) electrodes Cz and Pz are close to the scalp reference electrode so that the recordings from Cz (or Pz) reflecting the difference between two electrical potentials measured at the scalp reference electrode and Cz (or Pz) are rather small at most time points. (ii) the brain activity in the corrected Cz and Pz can be seen. (iii) Muscle artifacts in the referential Cz and Pz were reduced compared to that in the corrected Cz and Pz. The reference signal is the same as that in Figure 2A. We also note that considerable muscle artifacts can still be seen in bipolar Cz-Pz, illustrating that these artifacts are not from the reference signal. More importantly, we note that the amplitude of the reference signal R2 is larger than that of the corrected RTD5 and RTD6 and that of the raw Cz and Pz.

Figure 2(b) shows energy change of raw and corrected RTD5, bipolar RTD5-RTD6, and R2. One can clearly see all energy values of the raw RTD5 are much greater than that of the corrected RTD5. The reason lies in the fact that the reference signal R2 has larger amplitude than the corrected RTD5. Hence, the reference signal with high relative amplitude will increase energy values, and higher energy value of referential EEG cannot reflect smaller energy value of the corrected EEG and leads to misinterpretation of EEG.

Figure 2(c) shows energy change of raw and corrected Pz, bipolar Cz-Pz, and R2. One can clearly see energy values of the corrected Pz are much greater than that of the referential Pz at most time points. The reason lies in the fact that the reference signal R2 has larger amplitude than the referential Pz. We note that the amplitude of the reference signal R2 is a little smaller than that of the corrected Pz. In this case, the reference signal with small relative amplitude may decrease energy values. As a result, smaller energy value of referential EEG cannot reflect higher energy value of the corrected EEG and leads to misinterpretation of EEG.

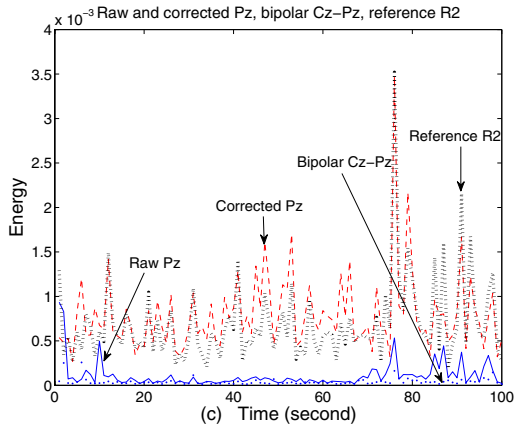
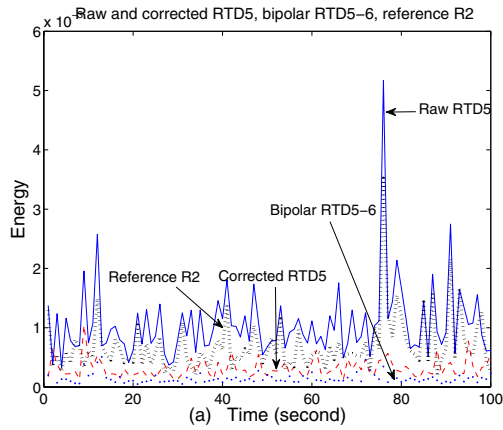
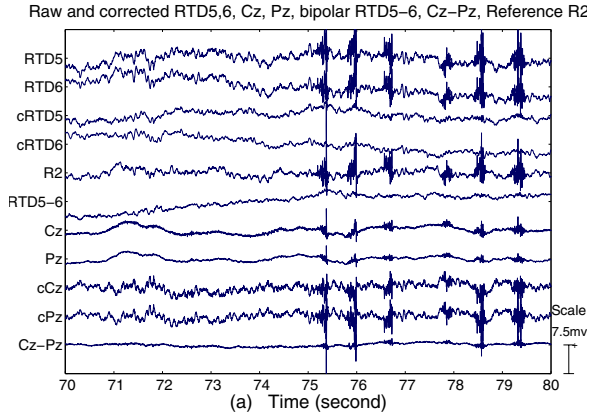


Fig. 2. (a) 10 seconds sample of EEGs (RTD5, RTD6, Cz, Pz, bipolar RTD5-RTD6, Cz-Pz) and the reference signal R2 where cRTD5, cRTD6, cCz and cPz are corrected RTD5, RTD6, Cz and Pz respectively after removing the reference signal R2. (b) Energy of raw and corrected RTD5, R2, and the bipolar RTD5-RTD6. (c) Energy of raw and corrected Pz, R2, and the bipolar Cz-Pz.

4 Discussion

In this study, we examined energy change of common referential, corrected, and bipolar EEG recorded from intracranial and scalp electrodes. We first obtained analytical expression for how energy of EEG depends on recording reference signal, and then using analytical and simulation approaches investigated the effect of the reference. We were able to show that (i) energy of the referential signal always monotonically increases as the amplitude of the reference signal increases from 0 to ∞ when the reference signal and the non-referential signal have negative correlation. (ii) Energy of the referential signal first decreases to some nonnegative value and then increases as the amplitude of the reference signal increases from 0 to ∞ when the reference signal and the non-referential signal have positive correlation. In general, the reference signal may decrease or increase energy values. But the reference signal with higher relative amplitude will surely increase energy values.

We investigated one patient undergoing evaluation for epilepsy surgery. In [1] we obtained the reference signal R2 by using the second method in [1] and explained why R2 is a “good” estimation of the real reference signal. After subtracting R2 from the referential signal we got corrected EEGs. Simulation results based on referential, corrected and bipolar EEGs showed that the reference signal may have a significant effect on energy values. For this particular patient, we found that the reference signal with smaller amplitude compared with scalp EEG may decrease energy values (see, the referential Pz of smaller values and corrected Pz of larger values in Figure 2-(c)). On the contrary, the reference signal with larger amplitude compared with iEEG may increase energy values (see, the referential RTD5 of larger values and corrected RTD5 of smaller values in Figure 2-(b)).

The commonly used bipolar EEG can remove the common reference. However, one should note that bipolar EEG will also remove all signals common to the two channels and not all signals common to the two electrodes are from the reference. Hence, a given bipolar montage will completely miss dipoles with certain locations and tangential orientations. Our simulation results from this patient demonstrated that bipolar EEG usually leads to small energy values and as a result cannot reflect real large energy values (see Figures 2-(b) and 2-(c)).

Acknowledgements. The work was supported by NIH 5K23NS047495, Epilepsy Therapy Development Program, and Mayo Clinic Discovery Translation Program.

References

1. Hu, S., Stead, S.M., Worrell, G.A.: Automatic identification and removal of scalp reference signal for intracranial EEGs based on independent component analysis. *IEEE Trans. on Biomedical Engineering* 54, 1560–1572 (2007)
2. Mormann, F., Andrzejak, R.G., Elger, C.E., Lehnertz, K.: Seizure prediction: the long and winding road. *Brain* 130, 314–333 (2007)
3. Esteller, R., Echauz, J., D’Alessandro, M., Worrell, G.A., Cranstoun, S., Vachtsevanos, G., Litt, B.: Continuous energy variation during the seizure cycle: towards an on-line accumulated energy. *Clin. Neurophysiol.* 116, 517–526 (2005)

4. Harrison, M.A., Frei, M.G., Osorio, I.: Accumulated energy revisited. *Clin. Neurophysiol.* 116, 527–531 (2005)
5. Gigola, S., Ortiz, F., D'Attellis, C.E., Silva, W., Kochen, S.: Prediction of epileptic seizures using accumulated energy in a multiresolution framework. *J. Neurosci. Methods* 138, 107–111 (2004)
6. Direito, B., Dourado, A., Vieira, M., Sales, F.: Combining energy and wavelet transform for epileptic seizure prediction in an advanced computational system. In: 2008 International Conference on BioMedical Engineering and Informatics, pp. 380–385 (2008)
7. Schiff, S.J.: Dangerous Phase. *Neuroinformatics* 3, 315–318 (2006)
8. Zaveri, H.P., Duckrow, R.B., Spencer, S.S.: On the use of bipolar montages for time-series analysis of intracranial electroencephalograms. *Clin. Neurophysiol.* 117, 2102–2108 (2006)
9. Hyvarinen, A., Oja, E.: Independent component analysis: algorithms and applications. *Neural Networks* 12, 411–430 (2000)

Multichannel Blind Deconvolution Using the Conjugate Gradient

Bin Xia^{1,2}

¹ Department of Electronic Engineering,
Shanghai Maritime University, Shanghai 200135, China

² College of Electronics and Information Engineering,
Tongji University, Shanghai 200065, China
binxia@cie.shmtu.edu.cn

Abstract. In this paper, we propose a conjugate gradient based algorithm for blind deconvolution. In general, blind deconvolution algorithms suffer from the speed of convergence. We make a further study of the geometrical structures on the manifold of finite impulse response (FIR) filters using lie group method. We derive the expressions of geodesic and parallel translation on the manifold of FIR filters. Using mutual information criteria, a feasible cost function is derived for blind deconvolution problem. Then we develop a conjugate gradient algorithm for multichannel blind deconvolution problem in finite impulse response (FIR) manifold. Computer simulations show the validity and effectiveness of this approach.

Keywords: Blind deconvolution, Natural gradient, Conjugate gradient.

1 Introduction

Blind deconvolution is to retrieve the independent source signals from sensor outputs by only using the sensor signals and certain knowledge on statistics of the source signals. A number of methods have been developed to deal with the blind deconvolution problem. These methods include the Bussgang algorithms [12], higher order statistical approach (HOS) [34] and the second-order statistics approach (SOS) [56].

The natural gradient, developed by Amari et al [7], and the relative gradient developed by Cardoso et al [8], improve learning efficiency in blind deconvolution [9]. Zhang et al [10,11] derived the natural gradient algorithm for training FIR filter in blind deconvolution problems. The natural gradient algorithms adjust the parameters of the demixing model in negative of the natural gradient direction. The cost function is decreasing most rapidly in this direction. But in practice, the cost function can not obtain the fastest convergence performance in the direction of the negative natural gradient. Zhang [12] analyzed the geometrical structure on manifold of nonsingular matrices and developed conjugate gradient algorithm for training the parameter on the nonsingular matrix manifold. In this paper, we investigate the geometrical structures on the FIR manifold. After introduced the geodesic and parallel translation, we develop conjugate gradient algorithm, which produces generally faster convergence than steepest descent direction, for blind deconvolution problem on FIR manifold.

2 Problem Formulation

Consider a convolutive multichannel mixing model, linear time-invariant (LTI) and non-causal systems of form

$$\mathbf{x}(k) = \mathbf{H}(z)\mathbf{s}(z), \quad (1)$$

where $\mathbf{H}(z) = \sum_{p=-\infty}^{\infty} \mathbf{H}_p z^{-p}$, z is the delay operator, \mathbf{H}_p is a $n \times n$ -dimensional matrix of mixing coefficients at time-lag p , which is called the impulse response at time p , $s(k) = [s_1(k), \dots, s_n(k)]^T$ is an n -dimensional vector of source signals with mutually independent components and $x(k) = [x_1(k), \dots, x_n(k)]^T$ is the vector of the sensor signals. The objective of multichannel blind deconvolution is to retrieve the source signals using only the sensor signals $x(k)$ and certain knowledge of the source signal distributions and statistics. We introduce a multichannel LTI systems as a demixing model

$$\mathbf{y}(k) = \mathbf{W}(z)\mathbf{x}(k), \quad (2)$$

where $\mathbf{W}(z) = \sum_{p=-\infty}^{\infty} \mathbf{W}_p z^{-p}$, $\mathbf{y}(k) = [y_1(k), \dots, y_n(k)]^T$ is an n -dimensional vector of the outputs and \mathbf{W}_p is an $n \times n$ -dimensional coefficient matrix at time-lag p .

In blind deconvolution problem, there exist scaling ambiguity and permutation ambiguity because some prior knowledge of source signals are unknown. We can rewrite (2) as

$$\mathbf{y}(k) = \mathbf{W}(z)\mathbf{x}(k) = \mathbf{W}(z)\mathbf{H}(z)\mathbf{s}(k) = \mathbf{P}\mathbf{A}\mathbf{D}(z)\mathbf{s}(k), \quad (3)$$

where $\mathbf{P} \in \mathbf{R}^{n \times n}$ is a permutation matrix, $\mathbf{A} \in \mathbf{R}^{n \times n}$ is a nonsingular diagonal scaling matrix. Then the global transfer function is defined by $\mathbf{G}(z) = \mathbf{W}(z)\mathbf{H}(z)$. The blind deconvolution task is to find a demixing filter $\mathbf{W}(z)$ such that

$$\mathbf{G}(z) = \mathbf{W}(z)\mathbf{H}(z) = \mathbf{P}\mathbf{A}\mathbf{D}(z), \quad (4)$$

where $\mathbf{D}(z) = \text{diag}\{z^{-d_1}, \dots, z^{-d_n}\}$.

In order to study the geometrical structure of FIR manifold, we first introduce some Lie group properties and then derive two important concepts in the next section.

3 Geometrical Structures

3.1 Lie Group

We introduce a Lie group to the manifold of FIR filters in order to define self-closed multiplication and inverse operations. In the manifold $\mathcal{M}(N)$, the operations of *multiplication* $*$ and *inverse* \dagger are defined as

$$\mathbf{A}(z) * \mathbf{B}(z) = [\mathbf{A}(z)\mathbf{B}(z)]_N \quad (5)$$

where $[\]_N$ is the truncating operator that any terms with orders higher than N are omitted.

$$\mathbf{B}^\dagger(z) = \sum_{p=0}^N \mathbf{B}_p^\dagger z^{-p} \quad (6)$$

where $\mathbf{B}_p^\dagger (p = 0, 1, \dots, N)$ are recurrently defined by $\mathbf{B}_0^\dagger = \mathbf{B}_0^{-1}, \mathbf{B}_1^\dagger = -\mathbf{B}_0^\dagger \mathbf{B}_1 \mathbf{B}_0^\dagger, \mathbf{B}_p^\dagger = -\sum_{q=1}^p \mathbf{B}_{p-q}^\dagger \mathbf{B}_q \mathbf{B}_0^\dagger, p = 1, \dots, N.$

For the sake of simplicity, we only give some properties of Lie Group here. The reader can directly refer [13] for more detail information.

Property 1:

$$\mathbf{A}(z) * (\mathbf{B}(z) * \mathbf{C}(z)) = (\mathbf{A}(z) * \mathbf{B}(z)) * \mathbf{C}(z), \tag{7}$$

Property 2:

$$\mathbf{B}(z) * \mathbf{B}^\dagger(z) = \mathbf{B}^\dagger(z) * \mathbf{B}(z) = \mathbf{E}(z) \tag{8}$$

Where $\mathbf{E}(z)$ is the identity filter. After introducing *multiplication* $*$ and *inverse* \dagger , it is obviously that both $\mathbf{B}(z) * \mathbf{C}(z)$ and $\mathbf{B}^\dagger(z)$ still remain in the manifold $\mathcal{M}(N)$ and the manifold $\mathcal{M}(N)$ forms a Lie group with the above operations.

Lie group has an important property that admits an invariant Riemannian metric. Let $\mathcal{T}_W(\mathcal{M}(N))$ be the tangent space of $\mathcal{M}(N)$ at $\mathbf{W}(z)$, and $\mathbf{P}(z), \mathbf{Q}(z) \in \mathcal{T}_W(\mathcal{M}(N))$ be the tangent filters. We introduce the inner product with respect to $\mathbf{W}(z)$ as $\langle \mathbf{P}(z), \mathbf{Q}(z) \rangle_{\mathbf{W}(z)}$ in the following way. Since $\mathcal{M}(N)$ is a Lie group, any $\mathbf{B}(z) \in \mathcal{M}(N)$ defines an onto-mapping: $\mathbf{W}(z) \rightarrow \mathbf{W}(z) * \mathbf{B}(z)$. The multiplication transformation maps a tangent filter $\mathbf{P}(z)$ at $\mathbf{W}(z)$ to a tangent filter $\mathbf{P}(z) * \mathbf{B}(z)$ at $\mathbf{W}(z) * \mathbf{B}(z)$. Therefore we can define a Riemannian metric on $\mathcal{M}(N)$, such that the right multiplication transformation is isometric, that is , it preserves the Riemannian metric on $\mathcal{M}(N)$,

$$\langle \mathbf{P}(z), \mathbf{Q}(z) \rangle_{\mathbf{W}(z)} = \langle \mathbf{P}(z) * \mathbf{B}(z), \mathbf{Q}(z) * \mathbf{B}(z) \rangle_{\mathbf{W}(z) * \mathbf{B}(z)} \tag{9}$$

for any $\mathbf{P}(z), \mathbf{Q}(z) \in \mathcal{T}_W(\mathcal{M}(N))$. If we define the inner product at the identity $\mathbf{E}(z)$ by

$$\langle \mathbf{P}(z), \mathbf{Q}(z) \rangle_{\mathbf{E}(z)} = \sum_{p=0}^N \text{tr}(\mathbf{P}_p \mathbf{Q}_p^T), \tag{10}$$

then $\langle \mathbf{P}(z), \mathbf{Q}(z) \rangle_{\mathbf{W}(z)}$ is automatically induced by

$$\langle \mathbf{P}(z), \mathbf{Q}(z) \rangle_{\mathbf{W}(z)} = \langle \mathbf{P}(z) * \mathbf{W}(z)^\dagger, \mathbf{Q}(z) * \mathbf{W}(z)^\dagger \rangle_{\mathbf{E}(z)} \tag{11}$$

From (11), we can calculate the Riemannian metric $\mathcal{G}(\mathbf{W})$. It is not necessary to derive the complexity expression of $\mathcal{G}(\mathbf{W})$ because the (11) already provides sufficient information to derive the natural gradient using geometrical approach.

3.2 Geodesics

Here we use the notation of $\mathbf{W}(z)$

$$\mathbf{W}(z) = \sum_{p=0}^N \mathbf{W}_p z^{-p}; \tag{12}$$

The geodesic could be obtained from the following calculus of variational problem

$$\text{dist}(\mathbf{W}_1(z), \mathbf{W}_2(z)) = \min_{\mathbf{W}_t(z)} \int_0^1 \langle \mathbf{W}'_t(z), \mathbf{W}'_t(z) \rangle_{\mathcal{W}_t(z)}^{1/2} dt, \tag{13}$$

subject to

$$\mathbf{W}_t(z) = \begin{cases} \mathbf{W}_1(z) & : t = 0 \\ \mathbf{W}_2(z) & : t = 1 \end{cases}$$

where $'$ is the notation for the derivative with respect to t . From the definition of the inner product of the tangent space of manifold $\mathcal{M}(n)$, we know that

$$\langle \mathbf{W}'_t(z), \mathbf{W}'_t(z) \rangle_{\mathbf{W}_t(z)} = \langle \mathbf{W}'_t(z) * \mathbf{W}_t^\dagger(z), \mathbf{W}'_t(z) * \mathbf{W}_t^\dagger(z) \rangle_{\mathbf{E}_t(z)} \quad (14)$$

We introduce a new differential variable,

$$d\mathbf{X}_t(z) = [d\mathbf{W}_t(z) * \mathbf{W}_t^\dagger(z)]_N \quad (15)$$

Substituting (14) and (15) into (13) we can derive

$$\mathbf{X}'_t(z) = \mathbf{C}(z), \quad (16)$$

where $\mathbf{C}(z)$ is a constant filter. From the relation in (15) and (16), we obtain the following equation

$$\frac{d\mathbf{W}(t)}{dt} = \mathbf{C}(z) * \mathbf{W}_t(z), \quad (17)$$

subject to $\mathbf{W}_t(z) = \mathbf{W}_1(z)|_{t=0}$, $\mathbf{W}_t(z) = \mathbf{W}_2(z)|_{t=1}$. Solving the equation, we obtain the definition of geodesic

$$\mathbf{W}_t(z) = (\mathbf{W}_2(z) * \mathbf{W}_1^\dagger(z))^t * \mathbf{W}_1(z). \quad (18)$$

If we change the condition in (17)

$$\frac{d\mathbf{W}(t)}{dt} = \mathbf{C}(z) * \mathbf{W}(t), \quad (19)$$

subject to $\mathbf{W}_t(z) = \mathbf{W}_1(z)|_{t=0}$, $\mathbf{W}_t(z) = \mathbf{W}_2(z)|_{t=1}$. The expression of geodesic is given by

$$\mathbf{W}_t(z) = [\exp(t\mathbf{X}_1(z) * \mathbf{W}_1(z)^\dagger)]_N * \mathbf{W}_1(z). \quad (20)$$

where $[\cdot]_N$ is the truncating operator such that any terms with orders higher than N are omitted.

3.3 Parallel Translation

In FIR manifold, if we move tangent filter along to another point like the moving in Euclidean space, it is impossible to get the tangent filter at the end. As shown in Fig. 1, \mathbf{X}_1 is the tangent filter at \mathbf{W}_1 and \mathbf{X}_2 is the tangent vector at \mathbf{W}_2 on the geodesic. If we simply move \mathbf{X}_1 to \mathbf{W}_2 along geodesic, we get the $\bar{\mathbf{X}}_1$ at \mathbf{W}_2 . It is obviously, the $\bar{\mathbf{X}}_1$ is not the tangent filter at \mathbf{W}_2 . In order to let the tangent filter to move along the geodesic, we introduce the concept of parallel translation. From the expression of geodesic (20), the tangent filter at $\mathbf{W}_t(z)$ on the geodesic is expressed by

$$\begin{aligned} \dot{\mathbf{W}}_t(z) &= \mathbf{X}_1(z) * \mathbf{W}_1^\dagger(z) * [\exp(t\mathbf{X}_1(z) * \mathbf{W}_1(z)^\dagger)]_N * \mathbf{W}_1(z) \\ &= \mathbf{X}_1(z) * \mathbf{W}_1^\dagger(z) * \mathbf{W}_t(z) \end{aligned} \quad (21)$$

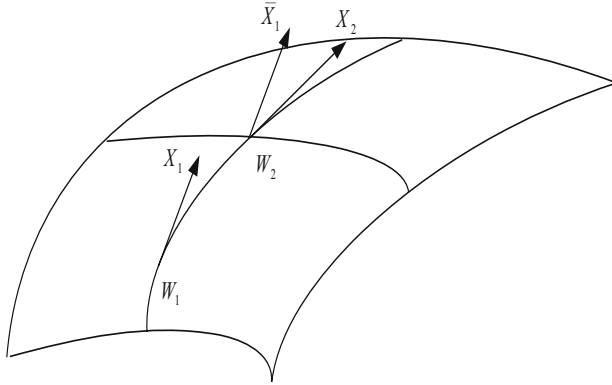


Fig. 1. Illustration of parallel translation

The tangent filter $\mathbf{X}_1 \in \mathcal{T}_{\mathbf{W}_1} \mathcal{M}(n)$ is translated into the tangent filter $\mathbf{X}_2 \in \mathcal{T}_{\mathbf{W}_2} \mathcal{M}(n)$ using Eq. (21)

$$\mathbf{X}_2 = \mathbf{X}_1 * \mathbf{W}_1^\dagger(z) * \mathbf{W}_2(z). \tag{22}$$

3.4 Natural Gradient

The gradient direction is not the steepest ascent direction for a cost function $l(\mathbf{W}(z))$ defined on Riemannian FIR manifold. We introduce the natural gradient on FIR manifold using a geometric approach. The ordinary gradient is denoted by

$$\nabla l(\mathbf{W}(z)) = \frac{\partial l(\mathbf{W}(z))}{\partial \mathbf{W}(z)} = \sum_{p=0}^N \frac{\partial l(\mathbf{W}(z))}{\partial \mathbf{W}_p} z^{-p}, \tag{23}$$

where

$$\frac{\partial l(\mathbf{W}(z))}{\partial \mathbf{W}_p} = \left(\frac{\partial l(\mathbf{W}(z))}{\partial \mathbf{W}_{p,ij}} \right)_{n \times n}, \quad p = 0, 1, \dots, N. \tag{24}$$

In order to derive the natural gradient on the manifold $\mathcal{M}(N)$, we introduce the following notations. The operator *vec* transforms a matrix $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n]$ to a vector $vec(\mathbf{A})$, defined by

$$vec(\mathbf{A}) = [\mathbf{a}_1^T, \mathbf{a}_2^T, \dots, \mathbf{a}_n^T]^T. \tag{25}$$

We further define the *vec* operator for a filter $\mathbf{P}(z)$ in $\mathcal{T}(\mathcal{M})(N)$ as

$$vec(\mathbf{P}(z)) = vec([\mathbf{P}_0, \mathbf{P}_1, \dots, \mathbf{P}_N]). \tag{26}$$

According to the definition of the natural gradient [9], we have

$$vec \tilde{\nabla} l(\mathbf{W}(z)) = \mathcal{G}(\mathbf{W})^{-1} vec(\nabla l(\mathbf{W}(z))). \tag{27}$$

We take the inner product with $\mathbf{P}(z)$ on the both sides of the above equation

$$\begin{aligned} & \langle \mathbf{P}(z), \nabla l(\mathbf{W}(z)) \rangle_{\mathbf{E}(z)} \\ &= \langle \text{vec}(\mathbf{P}(z)), \mathcal{G}(\mathbf{W}) \text{vec}(\tilde{\nabla} l(\mathbf{W}(z))) \rangle \\ &= \langle \mathbf{P}(z), \tilde{\nabla} l(\mathbf{W}(z)) \rangle_{\mathbf{W}(z)} \end{aligned} \quad (28)$$

There exists an geometrical interpretation: if we consider the filter $\mathbf{P}(z)$ as an element in $\mathcal{T}_{\mathbf{W}}(\mathcal{M}(N))$, then the inner product of $\mathbf{P}(z)$ and $\tilde{\nabla} l(\mathbf{W}(z))$ at $\mathbf{W}(z)$ is independent of $\mathbf{W}(z)$.

Combining (11) and (28), we can derive the natural gradient in the following way,

$$\begin{aligned} & \langle \text{vec}(\mathbf{P}(z)), \text{vec}(\nabla l(\mathbf{W}(z))) \rangle \\ &= \langle \text{vec}(\mathbf{P}(z)), \text{vec}(\tilde{\nabla} l(\mathbf{W}(z))) * \mathbf{W}^{-1}(z) \dots \\ & \quad * \mathbf{W}^{-T}(z^{-1}) \rangle, \end{aligned} \quad (29)$$

for any $\mathbf{P}(z)$ in $\mathcal{M}(N)$. Comparing the two sides of the above equation, we obtain

$$\tilde{\nabla} l(\mathbf{W}(z)) = \nabla l(\mathbf{W}(z)) * \mathbf{W}^T(z^{-1}) * \mathbf{W}(z). \quad (30)$$

Consider the differential $dl(\mathbf{W}(z))$ with respect to $\mathbf{X}(z)$ and $\mathbf{W}(z)$, respectively,

$$\begin{aligned} dl(\mathbf{W}(z)) &= \left\langle \frac{\partial l(\mathbf{W}(z))}{\partial \mathbf{X}(z)}, d\mathbf{X}(z) \right\rangle \\ &= \left\langle \frac{\partial l(\mathbf{W}(z))}{\partial \mathbf{W}(z)}, d\mathbf{W}(z) \right\rangle \\ &= \left\langle \frac{\partial l(\mathbf{W}(z))}{\partial \mathbf{W}(z)} * \mathbf{W}^T(z^{-1}), d\mathbf{X}(z) \right\rangle \end{aligned} \quad (31)$$

So we can obtain

$$\frac{\partial l(\mathbf{W}(z))}{\partial \mathbf{X}(z)} = \frac{\partial l(\mathbf{W}(z))}{\partial \mathbf{W}(z)} * \mathbf{W}^T(z^{-1}). \quad (32)$$

Substituting (32) into (30), we obtain the natural gradient

$$\tilde{\nabla} l(\mathbf{W}(z)) = \frac{\partial l(\mathbf{W}(z))}{\partial \mathbf{X}(z)} * \mathbf{W}(z). \quad (33)$$

3.5 Conjugate Gradient

In general, conjugate gradient algorithm can be described as follow: Given a initialization condition, we calculate the current search direction based on gradient method. The geodesic should be computed in current search direction and we take a line search to determine the optimal point by moving along the geodesic. The next search direction is combined with both current search direction and the natural gradient of the optimal point. In order to develop conjugate gradient for blind deconvolution problem, we should develop the natural gradient algorithm at first. The Kullback-Leibler Divergence

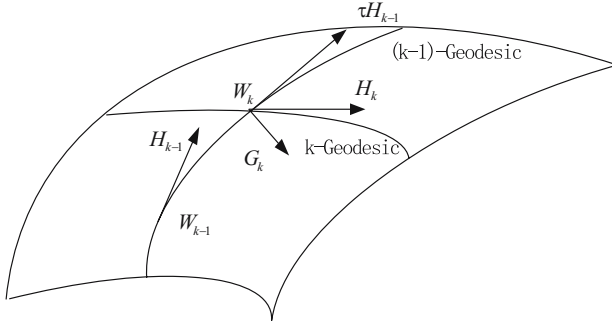


Fig. 2. Conjugate gradient in FIR manifold

has been used as a cost function for blind deconvolution [13] to measure the mutual independence of the output signals. They introduced the following simple cost function for blind deconvolution

$$l(\mathbf{y}, \mathbf{W}(z)) = -\log |\det(\mathbf{W}_0)| - \sum_{i=1}^n \log p_i(y_i). \tag{34}$$

where the output signals $y_i = \{y_i(k), k = 1, 2, \dots\}, i = 1, \dots, n$ as stochastic processes and $p_i(y_i(k))$ is the marginal probability density function of $y_i(k)$ for $i = 1, \dots, n$ and $k = 1, \dots, T$. The first term in the cost function is introduced to prevent the matrix \mathbf{W}_0 from being singular. The natural gradient of the cost function [34] is given by

$$\tilde{\nabla} l(\mathbf{y}, \mathbf{W}_p) = \sum_{q=0}^p (-\delta_{0,q} \mathbf{I} + \boldsymbol{\varphi}(\mathbf{y}(k)) \mathbf{y}^T(k-q)) \mathbf{W}_{p-q} \tag{35}$$

where $\boldsymbol{\varphi}(\mathbf{y}(k)) = (\varphi_1(y_1), \varphi_2(y_2), \dots, \varphi_n(y_n))^T$, and $\varphi_i(y_i) = -q'_i(y_i)/q_i(y_i)$ is the activation function of y_i . Using the natural gradient descent learning rule we present a novel learning algorithm as follows

$$\Delta \mathbf{W}_p = \eta \sum_{q=0}^p (\delta_{0,q} \mathbf{I} - \boldsymbol{\varphi}(\mathbf{y}(k)) \mathbf{y}^T(k-q)) \mathbf{W}_{p-q} \tag{36}$$

In Fig. 2, we suppose that $\mathbf{W}_{k-1}(z)$ is the $(k-1)$ -th approximate solution and $\mathbf{H}_{k-1}(z)$ is the current search direction. The geodesic can be calculated as

$$\mathbf{W}_{k,t}(z) = \left[\exp(t \mathbf{H}_{k-1}(z) * \mathbf{W}_{k-1}^\dagger(z)) \right]_N * \mathbf{W}_{k-1}(z), \tag{37}$$

A line search is then performed to determine the \mathbf{W}_k along the geodesic.

$$\mathbf{W}_k = \arg \min_t \{l(\mathbf{W}_{k-1,t}(z))\}. \tag{38}$$

The new search direction $\mathbf{H}_k(z)$ is a combination of the old search direction $\mathbf{H}_k(z)$ and the new natural gradient $\mathbf{G}_k(z) = \tilde{\nabla}l(\mathbf{y}, \mathbf{W}(z))$

$$\mathbf{H}_k(z) = \mathbf{G}_k(z) + \gamma_k \tau \mathbf{H}_{k-1}(z). \tag{39}$$

where γ_k is chosen such that the new direction is conjugate to the previous search direction and the $\tau \mathbf{H}_{k-1}(z)$ is the parallel translation of $\mathbf{H}_{k-1}(z)$. γ_k can be computed by [12]

$$\gamma_k = \frac{\langle \mathbf{G}_k(z) - \tau \mathbf{G}_{k-1}(z), \mathbf{G}_k(z) \rangle}{\tau \mathbf{G}_{k-1}(z), \tau \mathbf{G}_{k-1}(z)}. \tag{40}$$

The conjugate gradient algorithm search the minimum point along the geodesic in each iteration step, which is usually using less iteration times than the natural gradient algorithm.

4 Simulation

In this section, we propose simulation to illustrate the performance of proposed conjugate gradient based blind deconvolution algorithm. We compare proposed algorithm with natural gradient based algorithm to verify the convergence performance. We build a minimum phase mixing model using state-space method. The source signals are three independent i.i.d. signals uniformly distributed in range (-1, 1).

To remove the effect of a single numerical trial, we use the ensemble average of 50 trails. Fig. 3 illustrates the comparison results. It shows the proposed algorithm’s convergence speed is faster than natural gradient algorithm’s.

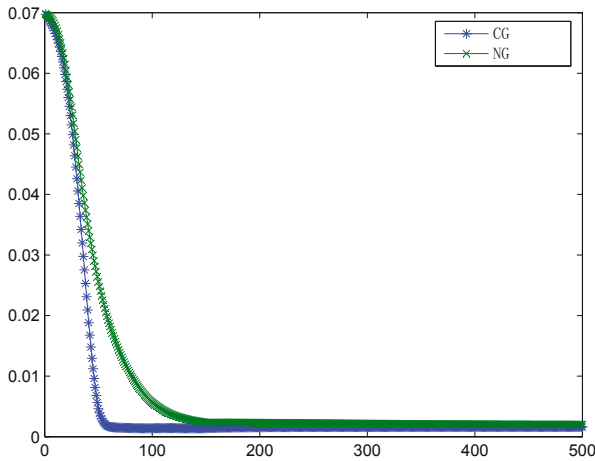


Fig. 3. Illustration of the performance of convergence, CG is Conjugate Gradient algorithm and NG is Natural Gradient algorithm

5 Conclusion

In this paper, we have investigated the geometrical structure of the FIR manifold and developed the conjugate gradient algorithm for blind deconvolution problem. Although the cost function decrease most rapidly along the negative of the natural gradient, this does not obtain the fastest convergence. Using the conjugate gradient algorithms, one-dimensional search for minimum point is along the geodesic, which produces generally faster convergence than steepest descent direction. The simulation shows the proposed algorithm obtain good performance of convergence.

Acknowledgments

The work was supported by Shanghai Maritime University project (No. 2008471).

References

1. Benveniste, A., Goursat, M., Ruget, G.: Robust Identification of a Nonminimum Phase System: Blind Adjustment of a Linear Equalizer in Data Communication. *IEEE Trans. Automatic Control* 25, 385–399 (1980)
2. Godard, D.N.: Self-Recovering Equalization and Carrier Tracking in Two-Dimensional Data Communication Systems. *IEEE Trans. Comm.* 28, 1867–1875 (1980)
3. Amari, S., Douglas, S., Cichocki, A., Yang, H.: Novel on-Line Algorithms for Blind Deconvolution Using Natural Gradient Approach. In: 11th IFAC Symposium on System Identification, Kitakyushu, Japan, pp. 1057–1062 (1997)
4. Bell, A.J., Sejnowski, T.J.: An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Computation* 7(6), 1129–1159 (1995)
5. Tong, L., Xu, G., Kailath, T.: Blind identification and equalization base on second-order statistics: A time domain approach. *IEEE Trans. Information Theory* 40, 340–349 (1994)
6. Tugnait, J.K., Huang, B.: Multistep Linear Predictors-Based Blind Identification and Equalization of Multiple-Input Multiple-Output Channels. *IEEE Trans. on Signal Processing* 48, 26–28 (2000)
7. Amari, S., Cichocki, A., Yang, H.H.: A New Learning Algorithm for Blind Signal Separation. In: Tesauro, G., Touretzky, D.S., Leen, T.K. (eds.) *Advances in Neural Information Processing Systems* 8, Cambridge, MA, pp. 757–763 (1996)
8. Cardoso, J., Laheld, B.: Equivariant Adaptive Source Separation. *IEEE Trans. on Signal Processing* 44(12), 3017–3030 (1996)
9. Amari, S.: Natural Gradient Works Efficiently in Learning. *Neural Computation* 10(2), 251–276 (1998)
10. Zhang, L.Q., Amari, S., Cichocki, A.: Semiparametric Model and Superefficiency in Blind Deconvolution. *Signal Processing* 81, 2535–2553 (2001)
11. Zhang, L.Q., Amari, S., Cichocki, A.: Geometrical Structures of Fir Manifold and Multichannel Blind Deconvolution. *Journal of VLIS for Signal Processing Systems* 31, 31–44 (2002)
12. Zhang, L.Q.: Geometric Structures and Unsupervised Learning on Manifold of Nonsingular Matrices. *Neural Computing* (submitted, 2004)
13. Zhang, L.Q., Amari, S., Cichocki, A.: Multichannel Blind Deconvolution of Non-Minimum Phase Systems Using Filter Decomposition. *IEEE Trans. Signal Processing* 52(5), 1430–1442 (2004)

An Improvement of HSMM-Based Speech Synthesis by Duration-Dependent State Transition Probabilities

Jing Tao and Wenju Liu

National Laboratory of Pattern Recognition, Institute of Automation,
Chinese, Academy of Sciences, Beijing 100190
{jtao, lwj}@nlpr.ia.ac.cn

Abstract. In this paper, we propose an improvement of hidden semi-Markov model (HSMM) based speech synthesis system by duration-dependent state transition probabilities. In traditional HMM algorithm, the probability of the duration of a state decreases exponentially with time, which does not provide an adequate representation of the temporal structure of speech. To overcome this limitation, HSMM, which models explicitly the state duration distribution, was proposed. However, there is still an inconsistency. Although HSMM has explicit state duration probability distributions, the state transition probabilities are duration-invariant. In this paper, we introduce duration-dependent state transition probabilities, which are able to characterize the timescale distortion at particular instant of an utterance more effectively, into HSMM based speech synthesis system. Correspondingly we improve forward-backward algorithm and re-derive parameter re-estimation formulae. Experimental results show that the proposed method improves the naturalness of the synthesized speech.

Keywords: Speech Synthesis, Duration-Dependent State Transition Probabilities, Forward-Backward Algorithm.

1 Introduction

A statistical parametric speech synthesis system based on hidden Markov models (HMMs) has made significant progress during the past decade. In this system, the spectrum, F0 and duration are modeled simultaneously in a unified HMM framework [1]. This method is able to synthesize highly intelligible and smooth speech. The most attractive part of this system is that its voice characteristics, speaking styles, or emotions can easily be modified by transforming HMM parameters using various techniques such as adaptation [2].

Although the HMM-based speech synthesis system has many advantages, the synthesized speech is still less natural compared with concatenation-based system. One of major drawback of the traditional use of HMM for speech synthesis is that the traditional HMM algorithms do not provide an adequate representation of the temporal structure of speech. This is because the probability of state occupancy decreases exponentially with time. To overcome this drawback, HSMM, which models explicitly the state duration probability distributions, was proposed [3]. However, there is

still an inconsistency. Although HSMM has explicit state duration probability distributions, the state transition probabilities are duration-invariant. In the present paper, we introduce duration-dependent state transition probabilities [4] into the hidden semi-Markov model based speech synthesis system. It describes the HMM is inhomogeneous as state transition probabilities are not irrelevant with time, and the transition probability from one state to another depend on the duration in the state. We call this new model a DDHSMM. This time varying transition probabilities which are able to characterize the timescale distortion at particular instant of an utterance more effectively and more in line with the characteristics of the time-domain of speech signal.

The rest of this paper is organized as follows. Section 2 briefly introduces the HMM-based speech synthesis system and the likelihood computation of HSMM. Section 3 describes the duration-dependent state transition probability, the likelihood computation of DDHSMM and derives its re-estimation formulae to construct a DDHSMM-based speech synthesis system. Experimental results are given in Section 4 and Section 5 concludes this paper.

2 HMM-Based Speech Synthesis System

2.1 HMM-Based Speech Synthesis System

Fig. 1 shows the overview of the current HMM-based speech synthesis system (HTS). It consists of training and synthesis parts. In the training part, both spectrum (mel-cepstral coefficients, their delta and delta-delta coefficients) and excitation (logarithmic fundamental frequencies ($\log F_0$) and its delta and delta-delta coefficients) parameters are extracted from a speech database. To model variable dimensional

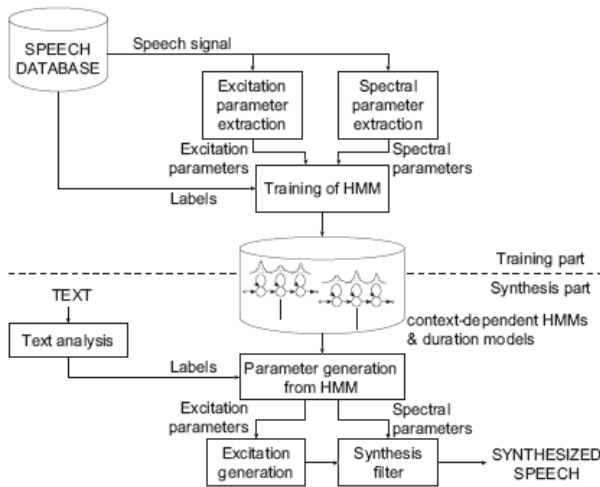


Fig. 1. An overview of the HMM-based speech synthesis system

parameter sequences such as $\log F_0$ with unvoiced regions properly, multi-space probability distributions (MSD) [5] are used. Each HMM has state duration probability density functions (PDFs) to model the temporal structure of speech [3], [6]. As a result, the system models spectrum, excitation, and durations simultaneously in a unified HMM framework [1].

In the synthesis part, an arbitrarily given text to be synthesized is converted to a context-dependent label sequence. Then a sentence HMM is constructed by concatenating context-dependent HMMs according to the label sequence. State durations of the sentence HMM are determined from the total length of speech and the state duration densities. According to the obtained state durations, a sequence of mel-cepstral coefficients is generated from the sentence HMM by using a speech parameter generation algorithm [7]. Finally, speech waveform is synthesized directly from the generated speech parameter vector sequence.

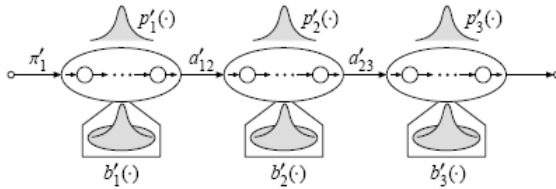


Fig. 2. An hidden semi-Markov model (HSMM) with 3-state left-to-right structures

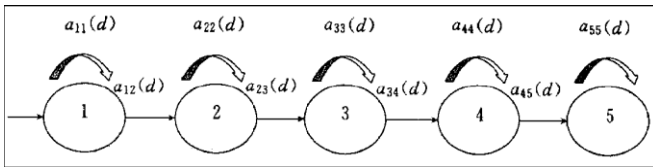


Fig. 3. DDHSMM topology

2.2 Likelihood Computation of the HSMM

HSMM can be considered as an HMM with explicit state duration probability distributions. We can compute the model likelihood of an HSMM λ illustrated in Fig. 2 for an observation vector sequence $O = (o_1, o_2, \dots, o_T)$ by the forward-backward algorithm [8]. We can compute partial forward likelihood $\alpha'_i(.)$ and partial backward likelihood $\beta'_i(.)$ recursively as follows:

$$\alpha'_i(i, d) = \begin{cases} \pi'_i b'_i(o_1) & \text{if } d = 1 \\ 0 & \text{otherwise} \end{cases}, \tag{1}$$

$$\alpha'_i(j, 1) = \sum_{\substack{i=1, \\ i \neq j}}^{N'} \sum_{d=1}^{D'} \alpha'_{i-1}(i, d) p'_i(d) a'_{ij} b'_j(o_t), \tag{2}$$

$$\alpha'_i(j, d) = \alpha'_{i-1}(j, d-1) b'_j(o_t), \tag{3}$$

$$\beta'_i(i, d) = p'_i(d) \sum_{\substack{j=1, \\ j \neq i}}^{N'} a'_{ij} b'_j(o_{t+1}) \beta'_{i+1}(j, 1) + b'_i(o_{t+1}) \beta'_{i+1}(i, d+1), \tag{4}$$

where $d, a'_{ij}, b'_j(o_t), N', \pi'_i, D'_i,$ and $p'_j(d)$ are a state duration, a state transition probability from state i to state j , an output probability of observation vector o_t from j , a total number of HSMM states, an initial state probability of state j , the maximum duration for state i , and a state duration probability of state j , respectively. From above equations, $p(o / \lambda')$ is computed as:

$$p(o / \lambda') = \sum_{i=1}^{N'} \sum_{d=1}^{D_i} \alpha'_i(i, d) \beta'_i(i, d). \tag{5}$$

3 Using Duration-Dependent State Transition Probabilities for HSMM-Based Speech Synthesis

From Eq. (2) and (4), we can see that, although HSMM has explicit state duration probability distribution $p'_j(d)$, the state transition probabilities a'_{ij} are duration-invariant. In this paper, we replace duration-invariant state transition probabilities with duration-dependent state transition probabilities.

3.1 Duration-Dependent State Transition Probability

The topology and state transitions of DDHSMM are illustrated in Fig. 3. We define duration-dependent state transition probabilities as follows:

$$a_{ij}(d) = P(q_{t+1} = j \mid q_t = i, d_t(i) = d), \quad 1 \leq i, j \leq N, 0 \leq d \leq D. \tag{6}$$

where N and D are the number of states and the maximum duration in any states, respectively. Eq. (6) represents the transition from state i to state j , given that the duration in state i at time t is $d_t(i) = d$.

3.2 Likelihood Computation of the DDHSMM Using an Improved Forward-Backward Algorithm

In order to introduce DDHSMM into speech synthesis, we use an improved forward-backward algorithm to calculate the probability of the observation sequence $O = (o_1, o_2, \dots, o_T)$ given the DDHSMM λ . In HTK [9], since the entry l and exit

states N of a HMM are non-emitting, we can set their duration as zero, that is $d_1 = d_N = 0$. For the forward probability, the initial conditions are established at time $t=1$ as follows

$$\alpha_1(1, d_1) = 1 \quad , \quad (7)$$

$$\alpha_1(j, 1) = \alpha_1(1, d_1) a_{1j}(d_1) b_j(o_1) \quad 1 < j < N \quad , \quad (8)$$

$$\alpha_1(N, d_N) = \sum_{i=2}^{N-1} \alpha_1(i, 1) p_i(1) a_{iN}(1) \quad , \quad (9)$$

All unspecified α values are zero. For time $1 < t \leq T$,

$$\alpha_t(j, 1) = \sum_{\substack{i=2 \\ i \neq j}}^{N-1} \sum_{d=1}^{D_i} \alpha_{t-1}(i, d) p_i(d) a_{ij}(d) b_j(o_t) \quad , \quad (10)$$

$$\alpha_t(j, d) = \alpha_{t-1}(j, d-1) b_j(o_t) \quad , \quad (11)$$

$$\alpha_t(N, d_N) = \sum_{i=2}^{N-1} \sum_{d=1}^{D_i} \alpha_t(i, d) p_i(d) a_{iN}(d) \quad . \quad (12)$$

where $a_{ij}(d)$ is the state transition probability from state i to state j , given that the duration in state i is d . $b_j(o_t)$ is the output probability of observation vector o_t from j and $p_i(d)$ is the state duration probability of state i . N is the number of states in DDHMM and D_i is the maximum duration in state i . For the backward probability, the initial conditions are set at time $t=T$ as follows

$$\beta_T(N, d_N) = 1 \quad , \quad (13)$$

$$\beta_T(i, d) = p_i(d) a_{iN}(d) \quad , \quad (14)$$

$$\beta_T(1, d_1) = \sum_{j=2}^{N-1} a_{1j}(d_1) b_j(o_T) \beta_T(j, 1) \quad , \quad (15)$$

Where once again, all unspecified β values are zero. For time $t < T$,

$$\beta_t(i, d) = p_i(d) \sum_{\substack{j=2 \\ j \neq i}}^{N-1} a_{ij}(d) b_j(o_{t+i}) \beta_{t+1}(j, 1) + b_i(o_{t+i}) \beta_{t+1}(i, d+1) \quad 1 < i < N \quad , \quad (16)$$

$$\beta_t(1, d_1) = \sum_{j=2}^{N-1} a_{1j}(d_1) b_j(o_t) \beta_t(j, 1) \quad . \quad (17)$$

The total probability can be computed by

$$P_r = P(O | \lambda) = \sum_{i=1}^N \sum_{d=1}^{D_i} \alpha_i(i, d) \beta_i(i, d) . \tag{18}$$

3.3 Parameter Re-estimation Formulae

In this section, we derive the parameter re-estimation formulae to construct a DDHSM-based speech synthesis system.

Using above forward and backward probabilities formulae, we can obtain the following variables for re-estimation of DDHSM parameters:

$$\xi_r(1, j, d_1) = \frac{1}{P_r} \alpha_r(1, d_1) a_{1j}(d_1) b_j(o_1) \beta_r(j, 1) , \tag{19}$$

$$\xi_r(i, j, d) = \begin{cases} \frac{1}{P_r} \alpha_r(i, d) p_i(d) a_{ij}(d) b_j(o_{t+1}) \beta_{t+1}(j, 1) & 2 \leq i, j < N \\ \frac{1}{P_r} \alpha_r(i, d) p_i(d) a_{iN}(d) \beta_r(N, d_N) & 2 \leq i < N \text{ and } j = N \end{cases} , \tag{20}$$

$$\gamma_r(i, d) = \frac{1}{P_r} \alpha_r(i, d) \beta_r(i, d) , \tag{21}$$

where $\gamma_r(i, d)$ is the probability in state i at time t with a duration of d and $\xi_r(i, j, d)$ is the probability of state transition from i to j at time $t+1$ after being in state i for a duration of d , given the model λ and observation O . The re-estimation formula of duration-dependent state transition probability $a_{ij}(d)$ is derived as follows:

$$a_{ij}(d) = \frac{\sum_{t=1}^T \xi_r(i, j, d)}{\sum_{t=1}^T \gamma_r(i, d)} . \tag{22}$$

In the DDHSM-based speech synthesis system, we still use single Gaussian distributions to model state duration probabilities. The re-estimation formulae of mean $\nu(i)$ and variance $\sigma(i)$ are derived as follows:

$$\nu(i) = \frac{\sum_{t=1}^T \sum_{d=1}^{D_t} \chi_t(i, d) d}{\sum_{t=1}^T \sum_{d=1}^{D_t} \chi_t(i, d)} , \tag{23}$$

$$\sigma(i) = \frac{\sum_{t=1}^T \sum_{d=1}^{D_t} \chi_t(i, d) d^2}{\sum_{t=1}^T \sum_{d=1}^{D_t} \chi_t(i, d)} - [v(i)]^2, \quad (24)$$

$$\chi_t(i, d) = \frac{1}{P_r} \alpha_t(i, d) p_i(d) \left[\sum_{\substack{j=2 \\ j \neq i}}^{N-1} a_{ij}(d) b_j(o_{t+1}) \beta_{t+1}(j, 1) \right. \\ \left. + a_{iN}(d) \beta_t(N, d_N) \right], \quad (25)$$

where $\chi_t(i, d)$ is the probability of state i at time t with a duration of d .

We use multi-stream structure with multi-space Gaussian distributions to model state output probability distributions [5]. Assuming that the g -th sub-space in the s -th stream is modeled by an n_{sg} -dimensional Gaussian distribution, the re-estimation formulae of the space weight $\hat{\omega}_{jsg}$, mean vector $\hat{\mu}_{jsg}$ and covariance matrix $\hat{\Sigma}_{jsg}$ are derived as follows:

$$\hat{\mu}_{jsg} = \frac{\sum_{t=1}^T \gamma_t(j, s, g) v(o_{st})}{\sum_{t=1}^T \gamma_t(j, s, g)}, \quad (26)$$

$$\hat{\Sigma}_{jsg} = \frac{\sum_{t=1}^T \gamma_t(j, s, g) [v(o_{st}) - \mu_{jsg}] [v(o_{st}) - \mu_{jsg}]^T}{\sum_{t=1}^T \gamma_t(j, s, g)}, \quad (27)$$

$$\hat{\omega}_{jsg} = \frac{\sum_{t=1}^T \gamma_t(j, s, g)}{\sum_{t=1}^T \sum_{d=1}^{D_t} \gamma_t(j, d)}, \quad (28)$$

$$\gamma_t(j, s, g) = \frac{1}{P_r} [\alpha_t(1, d_1) a_{1j}(d_1) \beta_t(j, 1) + \sum_{\substack{i=2 \\ i \neq j}}^{N-1} \sum_{d=1}^{D_i} \alpha_{t-1}(i, d) p_i(d) a_{ij}(d) \beta_t(j, 1) \\ + \sum_{d=2}^{D_j} \alpha_{t-1}(j, d-1) \beta_t(j, d)] \omega_{jsg} N(v(o_{st}); \mu_{jsg}, \Sigma_{jsg}) \prod_{k=1, k \neq s}^S b_{jk}(o_{kt}) \quad (29)$$

where $o_{st} = (X_{st}, x_{st})$, X_{st} is a set of space indexes, x_{st} is a continuous random variable, and $V(o_{st})$ is functions to extract x_{st} from o_{st} .

4 Experimental

4.1 Experimental Conditions

We use phonetically balance 3694 sentences from a speech database with a female speaker. Speech signals are sampled at 16 kHz and windowed by a 25ms Blackman window with a 10ms shift. Feature vector consists of 24-order mel-cepstral coefficients and pitch parameter vectors, including the zeroth coefficient, their delta and delta-delta coefficients. We use 5-state left-to-right HMMs with single diagonal Gaussian output distributions.

4.2 Experimental Results and Analysis

To evaluate the effectiveness of the proposed DDHSMM training, we use the spectral distortion distance:

$$E(A, B) = \frac{1}{N} \sum_{n=1}^N \sqrt{\frac{1}{M} \sum_{m=1}^M (S_A^{nm} - S_B^{nm})^2} \quad , \quad (30)$$

where N is the total number of frames, M is the dimensions of mel-cepstral coefficients, and S_A^{nm} is the value of m -th dimension in n -th frame. We generate 20 test sentences including in the training data using the DDHSMM-based system and HSMM-based system, respectively. Then we calculate the spectral distortion distance between generated spectra and natural spectra. Table 1 shows the results. We can see that the proposed method has lower spectral distortion distance. This is because DDHSMM contains more information of state transition, which is able to characterize the time-scale distortion at particular instant of an utterance more effectively.

Table 1. Spectral distortion distance compares

	HSMM-based	DDHSMM-based
SDD	0.0336	0.0320

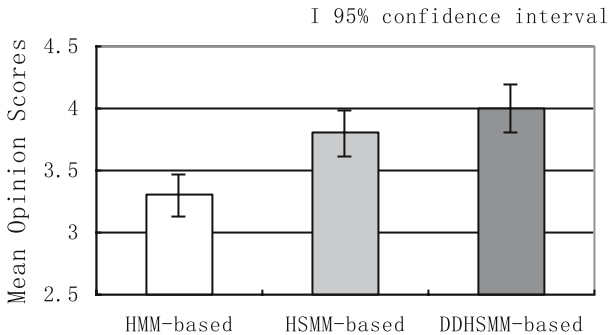


Fig. 4. The MOS results

We also use a mean opinion score (MOS) test to evaluate the quality of synthetic speech. We generate 40 test sentences which are not included in the training data, using the HMM-based, HSMM-based and DDHSMM-based system, respectively. Subjects are 10 persons, all of them are asked to provide a rating for the speech quality from 1.0 to 5.0. Fig. 4 shows the preference scores for the three systems. As we can see, the score of our system is higher than others.

5 Conclusion

In the present paper, we introduce duration-dependent state transition probabilities into the hidden semi-Markov model based speech synthesis system. Our method makes the HSMM more consistent and is able to characterize the timescale distortion at particular instant of an utterance more effectively. Experiment results show that our system can improve the quality of synthetic speech.

Acknowledgements. This work was supported in part by the China National Nature Science Foundation (No. 60675026, No. 60121302, No. 90820011), the 863 China National High Technology Development Projects (No. 20060101Z4073, No. 2006AA01Z194) and the National Grand Fundamental Research 973 Program of China (No. 2004CB318105).

References

1. Yoshimura, T., Tokuda, K., Masuko, T., Kobayashi, T., Kitamura, T.: Simultaneous Modeling of Spectrum, Pitch and Duration in HMM-Based Speech Synthesis. In: Proc. of Eurospeech, vol. 5, pp. 2347–2350 (1999)
2. Tamura, M., Masuko, T., Tokuda, K., Kobayashi, T.: Adaptation of Pitch and Spectrum for HMM-based Speech Synthesis Using MLLR. In: Proc. of ICASSP, pp. 805–808 (2001)
3. Zen, H., Tokuda, K., Masuko, T., Kobayashi, T., Kitamura, T.: Hidden Semi-Markov Model Based Speech Synthesis. In: Proc. of ICSLP, pp. 1185–1190 (2004)
4. Ramesh, P., Wilpon, J.G.: Modeling State Durations in Hidden Markov Models for Automatic Speech Recognition. In: Proc Int'l Conf. Acoustics, Speech, and Signal Processing, San Francisco, pp. 381–384 (1992)
5. Tokuda, K., Masuko, T., Miyazaki, N., Kobayashi, T.: Hidden Markov Models based on Multi-space Probability Distribution for Pitch Pattern Modeling. In: Proc. of ICASSP, pp. 229–232 (1999)
6. Yoshimura, T., Tokuda, K., Masuko, T., Kobayashi, T., Kitamura, T.: Duration Modeling for HMM-Based Speech Synthesis. In: Proc. ICSP, pp. 29–321 (1998)
7. Tokuda, K., Kobayashi, T., Imai, S.: Speech Parameter Generation from HMM Using Dynamic Features. In: Proc. of ICASSP, pp. 660–663 (1995)
8. Zen, H.: Implementing an HSMM-based Speech Synthesis System Using an Efficient Forward-Backward Algorithm. Technical Report of Nagoya Institute of Technology, TR-SP-0001 (2007)
9. Young, S., et al.: The HTK Book, Cambridge (2002)

Handprint Recognition: A Novel Biometric Technology

Guiyu Feng^{1,2}, Qi Zhao¹, Miyi Duan¹, Dewen Hu², and Yabin Hu³

¹ Beijing Graphics Research Institute, Chaoyang District, Beijing 100029, China
smartfgy@yahoo.com.cn

² College of Mechatronics and Automation,
National University of Defense Technology, Changsha 410073, China
dwhu@nudt.edu.cn

³ Central Laboratory of Information Technology,
China University of Geosciences, Wuhan 430074, China
wawahust@163.com

Abstract. In this paper, we present a novel biometric technology—handprint recognition. Handprint is obtained from the inner surface of a hand between the wrist and the top of the fingers, which contains the principal lines, wrinkles and ridges on the palm, finger and fingerprint. This paper discusses the advantages of this novel biometric: simple preprocessing and more distinctive features deployed. Furthermore, we make some elementary experiments on some essential aspects of this technology including preprocessing, feature representation and classifier design. The preliminary experimental results on the dataset sized of 50 persons are very encouraging, which suggests that more research work should be carried on this novel biometric.

Keywords: Pattern Recognition, Novel biometric, Handprint Recognition.

1 Introduction

System and information security is becoming increasingly important with the development of our human society. Personal identification is critical in many occasions such as e-commerce and access control. It is widely acknowledged that only biometric identifiers come close to actually identifying the persons rather than their possession or their exclusive knowledge [1].

Many researchers have carried research on various biometrics including fingerprint, voice, face and iris. Iris recognition can gain the highest recognition accuracy [2], however, the acquisition devices are expensive and may discomfort the users. In recent years, personal identification based on face and voice has been getting hot, but their performances are far from satisfactory [3]. As is well known that three biometrics can be extracted from the hand, i.e., fingerprint, hand geometry and palm-
print. Fingerprint-based personal identification drew considerable attention and has become a relatively ready solution to many end needs [4], however, workers and the elderly may not provide clear fingerprints because their problematic skins or physical work. Hand geometrical features such as finger width, length, and the thickness are adopted to represent extracted features, but these features frequently vary due to the

wearing of rings in fingers, besides, the width of some fingers may vary during pregnancy or illness, in addition, these features are no distinctive enough. The palmprint contains much more distinctive features than fingerprint and is protected by the hand, furthermore, the palmprint images are much easily to obtain, which attracted much research work on it [5-8, 10-13]. However, only the central part of the inner surface of the hand is considered to discriminate different persons, which does not take into account the information from other parts of the hand images.

In this paper, handprint recognition, as a novel biometric technology, is proposed. Handprint is obtained from the inner surface of a hand between the wrist and the top of the fingers, which contains the principal lines, wrinkles and ridges on the palm, finger and fingerprint. Hand features for identity verification gain the popularity from its user friendliness, environment flexibility and discriminating ability. People will not feel uneasy to have their hand images and prints taken for security purposes. More importantly, these hand features are stable and can uniquely represent each individual's identity. In addition, the handprint feature contains the traditional palmprint feature, fingerprint feature and at the same time, the hand geometry feature, which makes contribution to multi-modal biometrics fusion. Consequently, it is natural to develop a novel biometric technology using handprint features for security considerations.

Any biometric system can be divided into two modes: identification and verification. Our system can work in both modes. There are three key issues to be considered in developing a novel handprint biometric system:

- (1) **Handprint Acquisition:** How to obtain a good quality handprint image is the first important step. Through our experiments, it is found that an optical scanner is suitable for handprint data acquisition.
- (2) **Handprint Feature Representation:** Feature extraction attracted most efforts from biometrics researchers. We tempt to use PCA for handprint feature representation.
- (3) **Classifier Design:** After feature representation method is decided, the following important step should be classifier design. In the newly proposed handprint recognition system, a relatively simple nearest neighbor classifier is adopted, which is in L2 norm sense.

The rest of this paper is organized as follows: Section 2 provides a description of the handprint images acquisition device and the preprocessing of handprint images. A handprint feature extraction scheme and nearest neighbor classifier are detailed in Section 3. Section 4 reports the experimental results. Finally, the conclusion and future work are presented in Section 5.

2 Handprint Image Acquisition and Preprocessing

2.1 Handprint Images Acquisition

In our experimental environment, an optical scanner is used to capture the hand images. Here, the scanner that we use in our system is a color scanner which is a commercial product of AGFA Co. Volunteers are asked to put their right hands on the platform of scanner around the corner of a fixed object as shown in Figure 1. In this paper, we capture handprint images with 100 *dpi* resolutions. Handprint images are obtained with size of 650*813 in BMP format.

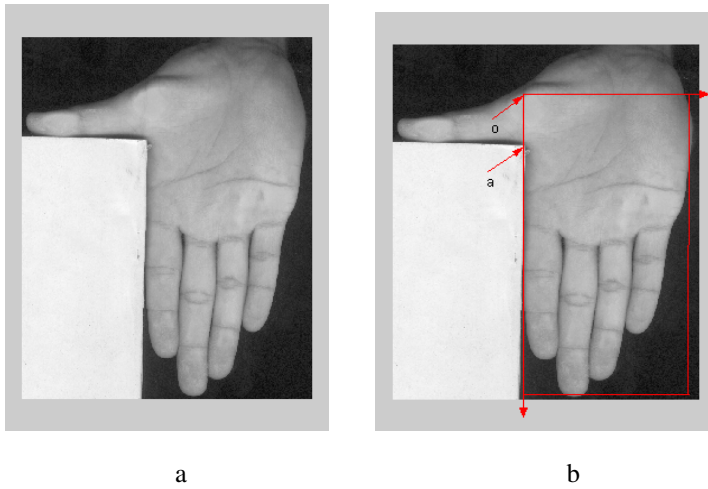


Fig. 1. The Preprocessing of Hand Images (a. Original image. b. The coordinate system.)

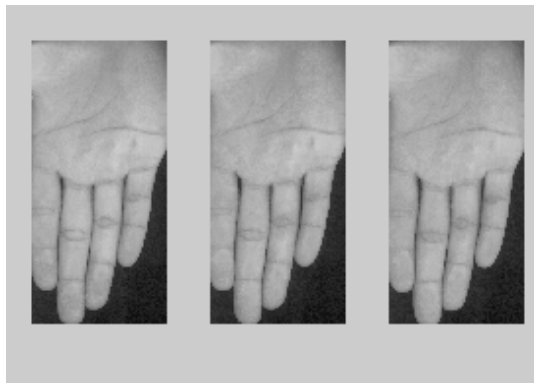


Fig. 2. Three handprint images from one subject

2.2 Preprocessing

Image preprocessing is usually the first and essential step in pattern recognition. It is necessary to define a coordinate system that is used to align different handprint images for further processing. To extract the nearly common part of a handprint for reliable feature measurements, a coordinate system is determined as shown in Fig. 1.b, firstly, the corner of the object, which is located manually and labeled as a in Fig. 1.b. Supposed the location of a is (x, y) , then o is decided as $(x, y-140)$, regard o as the new origin, the coordinate system parallel to the original one is defined. Fig. 2 shows three examples of the extracted handprint images.

3 Handprint Feature Representation and Classifier Design

Like palmprint, there are many approaches for handprint feature extraction, which could be based on structural features, statistical features or algebraic features [7]. However, structural features such as principal lines, wrinkles, delta points, minutiae, feature points and interesting points are difficult to be extracted, statistical features such as texture analysis tend to be not distinctive enough, PCA optimizes the transformation matrix by finding the largest variations in the original feature space, which is used as a typical algebraic feature and gains great success in both face recognition [9] and palmprint recognition [5], therefore, PCA is adopted for handprint feature analysis in our experiment.

3.1 Handprint Feature Representation Using PCA

Let the training samples of the handprints be x_1, x_2, \dots, x_M , where M is the number of images in the training set. The average image of the training set is defined by $\mu = \frac{1}{M} \sum_{i=1}^M x_i$. The difference between each image and the average image is given by $\varphi_i = x_i - \mu$. Then, we can obtain the covariance matrix of $\{x_i\}$ as follows:

$$C = \frac{1}{M} \sum_{i=1}^M (x_i - \mu)(x_i - \mu)^T = \frac{1}{M} XX^T \tag{1}$$

where the matrix $X = \{\varphi_1, \varphi_2 \dots \varphi_M\}$. It is evident that the matrix C can span an algebraic eigenspace and provide an optimal approximation for those training samples in terms of the mean-square error. It is well known that the following formula is satisfied for the matrix C :

$$C\mu_k = \lambda_k \mu_k \quad (k = 1, 2, \dots, M) \tag{2}$$

where μ_k refers to the eigenvector of matrix C , and λ_k is the correlative eigenvalue of matrix C . The significant eigenvectors μ_k with the largest associated eigenvalues are then selected to be the component of the eigenhands $U = \{\mu_k, k = 1, 2, \dots, M\}$, which can span M dimensional subspace of all possible images. A new image is transformed into its eigenspace by the following operation:

$$p_i = U^T (x_i - \mu) \quad (i = 1, 2, \dots, M) \tag{3}$$

3.2 Classifier Design

After feature representation method is decided, the following important step should be classifier design; the well known nearest neighbor classifier in L2 norm sense is used to give the final identity decision. Suppose there are K persons in the database, and $p_i (i = 1, 2, \dots, K)$ is the corresponding person's feature template, then for identification case:

$$x \text{ is person } \hat{i} \text{ if } \|x - p_i\|_2 = \min_{j \in K} \|p - x_j\|_2 \quad (4)$$

for verification case:

$$\begin{aligned} x \text{ is person } \hat{i} \text{ if } \|x - p_i\|_2 &\leq \text{threshold} \\ x \text{ is not person } \hat{i}, &\text{ otherwise.} \end{aligned} \quad (5)$$

4 Experimental Results and Analysis

4.1 Handprint Database

We collected handprint images from 50 persons using the capture device described in Section 2.1, and the session intervals were from one week to two months. The subjects consisted of volunteers from the students and teachers at the lab of NUDT. In this database, 47 people are male, and the age distribution of the subjects is from 20 to 40, every subject was asked to provide 3 images with the right hand at either of the two sessions, therefore, there are totally 300 handprint images, for some examples after preprocessing, see Fig. 2.

4.2 Experiment Results and Analysis

In the experiment, all the handprint images are divided into two parts, three images of each hand from the first session are regarded as the training set, and the left comprise the test set.

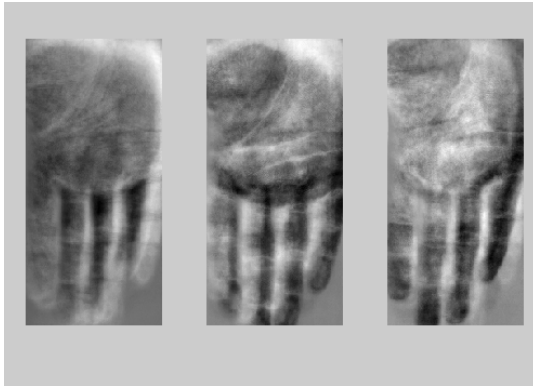


Fig. 3. Three eigenhands with largest eigenvalues

After histogram equalization on the handprint images, PCA is used to extract the features, the eigenvectors are called eigenhands, for visualization, see Fig.3, then the nearest neighbor strategy is adopted to give the final decision. Figure 4 gives the identification performance results. From Fig. 4, we can find that the recognition accuracy rate could be as high as 94.67%, from which we can see that the handprint images contain enough distinctive features to distinguish from each other. In Figure 5, the

verification performance indicator—ROC curve is given; the Equal Error Rate (EER) could be as low as 4.66%. Concerning the relatively simple preprocessing step and the feature representation method, it is certain that there is still much space for the performance of this biometric system to improve. The findings are not surprising because the handprint contains the palmprint, finger and as well as fingerprint. It should also be pointed out that the image resolution is only 100dpi, which is acceptable and will not make the storage space unavailable. If we increase the image resolution and make the preprocessing more precise, the performance should be more significant, and then we can safely draw the conclusion that handprint recognition, as a novel biometric solution strategy, should be paid more attention. We tentatively give a comparison of the four biometric technologies based on hand in Table 1, from which we can see the advantages of our newly proposed handprint biometric.

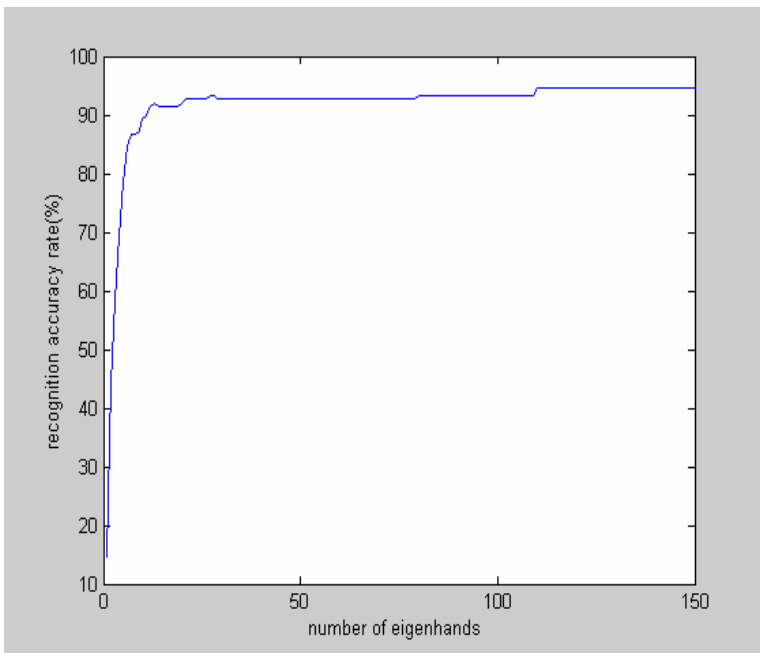


Fig. 4. The identification performance results of the handprint system

Table 1. Comparison of four biometric technologies based on hand

	Hand geometry	Fingerprint	Palmprint	Handprint
Performance	Low	High	High	High
Features	Inadequate	Adequate	Adequate	Most
Application	Authentication	Identification/Authentication	Identification/Authentication	Identification/Authentication

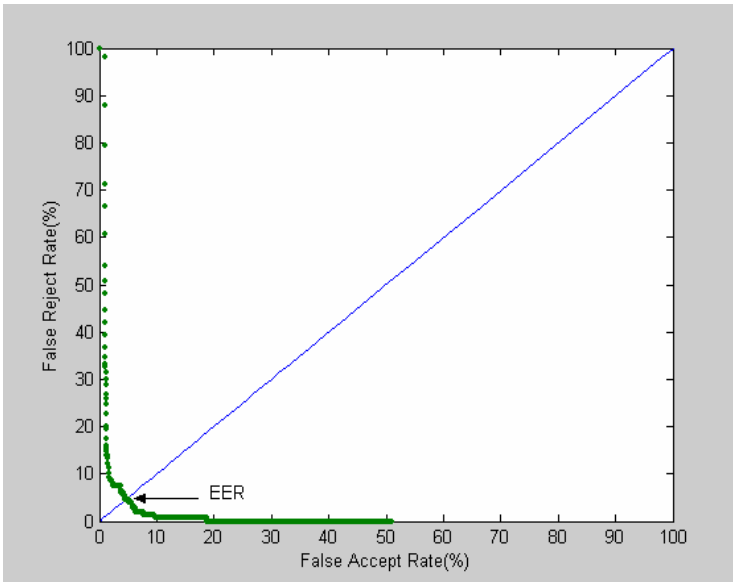


Fig. 5. The ROC curves of the handprint system

5 Conclusions and Future Work

In this paper, handprint recognition, as a novel biometric technology, is proposed. Compared with other biometrics, firstly, the preprocessing, i.e., the handprint images are quite easy to extract, secondly, handprint includes three parts: palmprint, finger and fingerprint, which contains more features than sole palmprint or fingerprint. Through experiments, it is found that with as low as 100 dpi resolutions, distinctive features can still be extracted from handprint images, which makes the storage space acceptable. The performance results of the proposed handprint system on the dataset sized of 50 persons are encouraging. As far as the relatively simple preprocessing step and the feature representation method are concerned, it is certain that there is still much room for the performance to improve. In addition, the handprint feature contains the traditional palmprint feature, fingerprint feature and at the same time, the hand geometry feature, which makes contribution to multi-modal biometrics fusion like [9]. In conclusion, our work suggests that more research work should be carried on this novel biometric. Our future work would include a more efficient handprint image acquisition device, which can capture handprint images with good quality and be processed by computers in real time; we would also enlarge the size of our handprint database. Handprint representation such as structural features and statistical features would be another research direction. Furthermore, other potential problems such as wearing of rings; width change of fingers during pregnancy or illness should also be investigated. Finally the handprint consists of the palm and the finger part, the relative role of these two parts is to be determined. Our ultimate goal is to develop a real time friendly handprint recognition system with good performance for access control purposes.

Acknowledgement. This work is partially supported by the National Natural Science Foundation of China (60675005, 60835005), 863 Program of China (2006AA01Z193) and the PLA General Armament Department (513150802).

References

1. Jain, A., Bolle, R., Pankanti, S. (eds.): *Biometrics: Personal Identification in Networked Society*. Kluwer Academic, Dordrecht (1999)
2. Ma, L., Tan, T., Wang, Y., Zhang, D.: Personal Identification Based on Iris Texture Analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 25, 1519–1533 (2003)
3. Pankanti, S., Bolle, R.M., Jain, A.: *Biometrics: the Future of Identification*. *IEEE Comp.* 33, 46–49 (2000)
4. Lee, H.L., Gaensslen, R.E. (eds.): *Advances in Fingerprint Technology*, 2nd edn. CRC Press, Boca Raton (2001)
5. Zhang, D., Shu, W.: Two Novel Characteristics in Palmprint Verification: Datum Point Invariance and Line Feature Matching. *Pattern Recognition* 32, 691–702 (1999)
6. Lu, G.M., Zhang, D., Wang, K.Q.: Palmprint Recognition Using Eigenpalms Features. *Pattern Recognition Letters* 24, 1463–1467 (2003)
7. Dong, K., Feng, G., Hu, D.: Digital Curvelet Transform for Palmprint Recognition. In: Li, S.Z., Lai, J.-H., Tan, T., Feng, G.-C., Wang, Y. (eds.) *SINOBIOMETRICS 2004*. LNCS, vol. 3338, pp. 639–645. Springer, Heidelberg (2004)
8. Zhang, D., Kong, W.K., You, J., Wong, M.: Online Palmprint Identification. *IEEE Trans. Pattern Anal. Mach. Intell.* 25, 1041–1050 (2003)
9. Feng, G., Dong, K., Hu, D., Zhang, D.: When Faces Are Combined with Palmprints: A Novel Biometric Fusion Strategy. In: Zhang, D., Jain, A.K. (eds.) *ICBA 2004*. LNCS, vol. 3072, pp. 701–707. Springer, Heidelberg (2004)
10. Hu, D., Feng, G., Zhou, Z.: Two-Dimensional Locality Preserving Projections (2DLPP) with its Application to Palmprint Recognition. *Pattern Recognition* 40, 339–342 (2007)
11. Feng, G., Hu, D., Zhou, Z.: A Direct Locality Preserving Projections Algorithm for Image Recognition. *Neural Processing Letters* 27, 247–255 (2008)
12. Feng, G., Zhang, D., Yang, J., Hu, D.: A Theoretical Framework for Matrix-Based Feature Extraction Algorithms with its Applications to Image Recognition. *International Journal of Image and Graphics* 8, 1–23 (2008)
13. Feng, G., Hu, D., Zhang, D., Zhou, Z.: An Alternative Formulation of Kernel LPP with Application to Image Recognition. *Neurocomputing* 69, 1733–1738 (2006)

Single Trial Evoked Potentials Estimation by Using Wavelet Enhanced Principal Component Analysis Method

Ling Zou^{1,2,3}, Zhenghua Ma¹, Shuyue Chen¹, Suolan Liu¹, and Renlai Zhou^{1,2,3}

¹ School of Information Science & Engineering, Jiangsu Polytechnic University,
Changzhou 213164, China

² State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University,
Beijing 100875, China

³ Beijing Key Lab of Applied Experimental Psychology, Beijing 100875, China
{Ling.Zou, Zhenghua.Ma, Shuyue.Chen, Suolan.Liu, Renlai.Zhou,
rlzhou}@bnu.edu.cn

Abstract. In this paper we present a new wavelet denoising (WD) enhanced principal component analysis (PCA) method (wPCA) to reduce the number of trials required for the efficient extraction of brain event related potentials (ERPs). First, the ERPs are extracted with wavelet transform, giving us an enhanced version of the raw data. Next, the principal components (PCs) with most of the total variance are considered to be part of the ERP subspace. Lastly, the ERPs are reconstructed from the selected PCs. Simulation and experimental results show that the wPCA method provides better performance than either WD or PCA method.

Keywords: Event related potentials, Wavelet transform, Principal component analysis, Single-trial extraction.

1 Introduction

In recent years, event related potentials (ERPs) analysis has become very useful for neuropsychological studies and clinical procedures [1-3]. The most common way to visualize ERPs has been to take an average over time locked single-trial measurements. The implicit assumption in the averaging is that the task-related cognitive process does not vary much in timing from trial to trial. However, it has been evident for a few decades that in many cases this assumption is not valid. The observation of variation in the parameters of ERPs permit the dynamic assessment of changes in cognitive state. Thus, the goal in the analysis of ERPs currently is the estimation of single potentials, which is called single-trial extraction. Several techniques have been proposed to improve the visualization of ERPs from the background electroencephalogram (EEG) with various successes [2-4].

Among these techniques, the wavelet transform (WT) method is especially promising for its optimal resolution both in the time and in the frequency domain. WT is an efficient tool for multiresolution analysis of non-stationary and fast transient signals.

These properties make it especially suitable to the study of neurophysiologic signals. Numerous WT applications in biosignal analysis have been proposed, including for the attempt of single-trial-ERP analysis [2, 5-7].

Principal component analysis (PCA) has been extensively used in feature extraction to reduce the dimensionality of the original data by a linear transformation. PCA extracts dominant features (principal components, PCs) from a set of multivariate data. The dominant features retain most of the information, both in the sense of maximum variance of the features and in the sense of minimum reconstruction error. PCA has been widely used in medical applications [8-10].

In this paper, we propose a new method to reduce the number of trials required for the efficient extraction of brain event related potentials (ERPs): wavelet denoising (WD) enhanced principal component analysis (PCA) method (wPCA). We first obtain an enhanced version of the raw data by WT denoising, then the principal components (PCs) with most (80%) of the total variance are considered to be part of the ERP subspace. Lastly, the ERPs are reconstructed from the selected PCs. Simulation and experimental results show that the wPCA method provided better capability than either WD or PCA method.

2 Methods

Multiple trials of observed ERPs can be modeled as

$$x_i(n) = s(n) + v_i(n) + z_i(n) \quad i = 1, 2, \dots, L; \quad 0 \leq n \leq N-1 \quad (1)$$

Where $s(n)$ are ERP components. The background neural activity is simulated as a mixture of colored noise $v_i(n)$ and Gaussian noise $z_i(n)$, which varies over trials. The objective is to extract the ERP signals $s(n)$ from the given L trials.

2.1 Wavelet Denoising (WD)

A classical solution for noise removal from non-stationary signals is WD. The basic principle is: the decomposition of a noisy signal on a wavelet basis (discrete wavelet transform, DWT) has the property to “concentrate” the informative signal in few wavelet coefficients having large absolute values without modifying the noise random distribution. After transformation the noise coefficients have small values, in contrast to the informative signal (normal or pathologic neural activity and artifacts). Therefore, denoising can be achieved by thresholding the wavelet coefficients.

Consider the i -th noisy mixture observation from (1),

$$x_i(n) = s(n) + v_i(n) + z_i(n) = s(n) + m(n) \quad (2)$$

Let W and W^{-1} be the forward and inverse DWT operators. WD can be performed for a given mixed signal x_i according to the following process:

$$w_i = W(x_i) \quad (3)$$

$$\hat{w}_i = T(w_i, \lambda) \quad (4)$$

$$\hat{c} = W^{-1} \left(\hat{w}_i \right) \tag{5}$$

Where w_i is the wavelet coefficient vector, $T(w_i, \lambda)$ is the thresholding operator with threshold λ , \hat{w}_i is the wavelet coefficients after thresholding and \hat{c}_i is the denoised signal.

The main problem is computing the threshold. There are four classical threshold deviation methods, including universal threshold, SURE threshold, hybrid threshold, and minimax threshold. In the ERPs study, not losing information that is potentially useful for medical diagnosis is of great importance. Moreover, because in ERPs the signal to noise ratio is low, the wavelet coefficients of neuronal signals can have small values compared to noise. Therefore low thresholding would be more appropriate. The hybrid thresholding method is used to determine the threshold value in the present study. After threshold selection, the thresholding process is accomplished by transforming the preserved data into a noise-reduced signal by the hard or soft transformation expressed. The soft-thresholding rule is used in the present study because the soft-thresholding method has a better mathematical characteristic over the hard thresholding. The choice of the wavelet type is also a practical issue in WD. In this study, we choose the Daubechies wavelets as the basic wavelet functions for their simplicity and general purpose applicability in a variety of time-frequency representation problems [11].

2.2 Principal Component Analysis (PCA)

Principal Component Analysis is one of the best known techniques in multivariate analysis. It is widely used in signal processing, statistical analysis, neural computing, financial prediction, etc. The main idea of PCA is to transform the data set to a new set of variables, which is called the principal components (PCs), which is arranged from the most variant basis to the least variant basis. Therefore, the first few principal components will retain most of the variation present in the original data set [12].

The PCA method is as follows.

Step 1: the ERP data X is arranged into an LxN matrix, where L is the number of the trials and N is the number of data samples in each trial.

$$X = [x_1 \quad x_2 \quad \dots \quad x_L]^T \tag{6}$$

x_1, x_2, \dots, x_n are the measurement vectors which represent the data from the respective trial.

Step 2: the covariance of matrix R is computed using

$$R = E[XX^T] \tag{7}$$

Step 3; compute V and D, where V is the orthogonal matrix of eigenvectors of R and D is the diagonal matrix of its eigenvalues,

$$D = \text{diag}(d_1, \dots, d_L) \quad (8)$$

Step 4: the principal components (PCs) are computed using

$$Y = V^T X \quad (9)$$

Step 5: the n^{th} diagonal value of D is the variance of X along the n^{th} PC, and the PCs with most (for instance 80-90%) of the total variance are selected to be part of the signal subspace, while the rest are considered to be part of the noise.

Step 6: the signal (without noise) is reconstructed from the selected PCs using

$$\hat{X} = \hat{V} \hat{Y} \quad (10)$$

Where \hat{V} and \hat{Y} are the eigenvectors and PCs with most of the total variance.

2.3 Wavelet Enhanced PCA (wPCA) Method

Among the several techniques that have been used to obtain enhanced single-trial ERPs, the wavelet technique has recently received more and more attention. However, this methodology is effective only in filtering white noise while the background EEG is the main noise in the measurements of ERP and the EEG is highly colored. Therefore it is difficult to extract ERP components from EEG records by the wavelet denoising technique alone. It should be pointed out that the PCA method provides better performance in the case of relatively high SNR. However, the SNR in experimental measurements is relatively low, and the conventional PCA method alone would not provide satisfied results.

Wavelets can be used to pre-process data in order to better locate and identify significant events [13]. PCA is a statistical process for feature extraction by reducing the data dimensionality using orthogonal bases. Combining this type of data pre-processing with multivariate statistics can generate useful insights into the problem of data analysis and data interpretation.

For the above reasons, we introduce the wavelet enhanced PCA (wPCA) method to reduce the number of trials required for the efficient extraction of brain ERPs. The method is illustrated as follows:

First, the ERPs are extracted by WD and we obtain an enhanced version X_{WT} of the raw data X . For a single-trial signal $x[n]$ we use Daubechies-6 wavelets, 5-level decomposition, universal thresholding and soft transformation. The WD method is applied to all the trials and we get the pre-denoised signal X_{WT} .

$$X_{WT} = [x_{WT1} \quad x_{WT2} \quad \dots \quad x_{WTL}]^T \quad (11)$$

Next, the PCA method is applied to the data X_{WT} and the PCs with 80% of the total variance are considered to be part of the ERPs subspace. Then the ERPs are reconstructed from the selected PCs.

3 Results

In this section, we provide simulated and experimental examples to compare the single-trial evoked potential estimation performance of the WD, PCA and the wPCA approaches.

3.1 Simulation Results

In this simulation, one simulated model for $s(n)$ [14] is used with sampling frequency 500Hz ($N=1024$), as shown in Fig. 1(a). The following autoregressive model is used for the colored noise $v_i(n)$ [15]:

$$v(n) = 1.5080v(n-1) - 0.1587v(n-2) - 0.3109v(n-3) - 0.0510v(n-4) + u(n) \quad (12)$$

Where $u(n)$ is zero-mean white Gaussian noise. We generate 15 trials of simulated data. Each trial contains 1024 data samples.

In the example, we have single channel with 15 trials, the latency of the ERP signal in the synthetic single trials is uniformly distributed within a predefined minimum and maximum value arbitrarily set to 100 to 400 ms, respectively. Thus, the corresponding latency range (jitter) could vary between 0 and 300ms.

Several merit measures are employed to assess the performance of different approaches. The first measure is the RMSE between the true ERP $s(n)$, and the estimated ERP $\hat{s}(n)$. The RMSE is computed as

$$RMSE = \left\{ \frac{1}{N} \sum_{n=1}^N \left(s(n) - \hat{s}(n) \right)^2 \right\}^{1/2} \quad (13)$$

The second measure is the SNR, which is computed as

$$SNR = 10 \log_{10} \left\{ \frac{\sum_{n=1}^N s^2(n)}{\sum_{n=1}^N \left(s(n) - \hat{s}(n) \right)^2} \right\} \text{ (db)} \quad (14)$$

In the simulation study, the 15 trials contain the same ERP. Fig. 1(b) corresponds to a single trial of the noisy raw signal. Fig.1(c)-(e) show the ERP estimates for the single trial obtained via the WD, PCA and wPCA approaches, respectively. The total SNR in the noisy raw signal is -2.531 db, while the total SNR in the estimated ERP is 6.796 db with WD, 1.88 db with PCA and 11.447 db with the wPCA method. From the figure we can see that the wPCA method has much more smooth estimates than the other methods. Smoothness is an expected property of the ERPs based on the reliable evoked potential estimation results obtained by ensemble averaging over a large number of trials. Table 1 compares the WD, PCA and the wPCA methods in terms of the mean and standard deviation (SD) of the RMSEs and SNRs of the ERP estimates for all trials. It is obvious from looking at the table that the wPCA approach obtains the most accurate estimates among all approaches. In addition, the WD method outperforms the conventional PCA method.

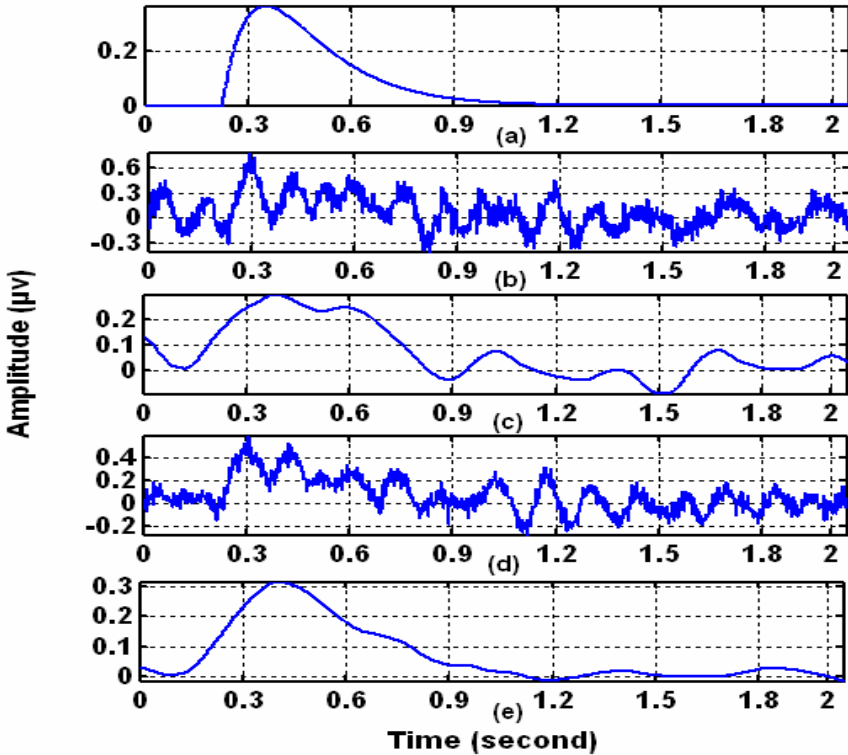


Fig. 1. Simulated results for a single trial from 15 trials in the presence of ERP variation: (a) simulated evoked potential; (b) noisy signal; (c) wavelet denoised signal; (d) PCA denoised signal; (e) the enhanced ERP signal using the wPCA approach

Table 1. Comparison of the different approaches for 15 trials in the presence of ERP variation

Approach	WD	PCA	wPCA
Mean of RMSEs	0.0620	0.1438	0.0386
SD of RMSEs	0.0065	0.0291	0.0069
Mean of SNRs (dB)	6.0370	-1.1547	10.2355
SD of SNRs (dB)	0.8982	1.7525	1.5258

3.2 Experimental Results

All experimental signals used in this paper are obtained from the emotion cognitive experiment at Beijing Normal University as we reported before [16]. Our previous studies have demonstrated that all the subjects show significantly greater P300 and slow waves amplitudes at medial-inferior and posterior electrode sites for pleasant and unpleasant pictures than for neutral pictures [16]. Here, we test the wPCA method by using the signals at the PO8 electrode of one subject. The performance of the WD, PCA and wPCA approaches on actual data is shown in Fig. 2. Fig. 2(a) corresponds to

a single trial of the “noisy” VEP signal from a subject under unpleasant pictures stimuli at the PO8 electrode site. Fig. 2(b)-(d) show the evoked potential estimates for the single trial obtained via the WD, PCA and wPCA methods, respectively. We see that the wavelet-based method has much more smooth estimates than the PCA method, and the wPCA method performs best.

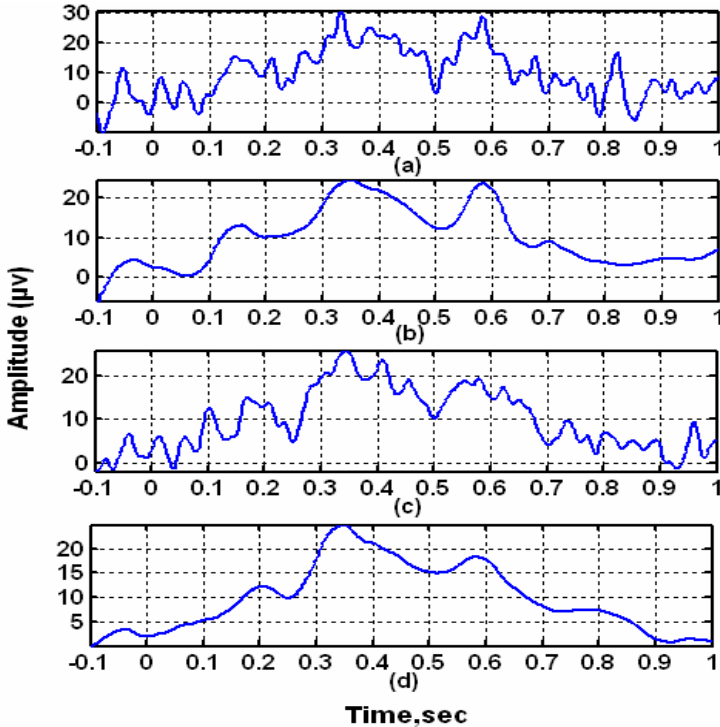


Fig. 2. Experimental results for a single trial of VEP: (a) a noisy signal; (b) wavelet denoised signal; (c) PCA denoised signal; (d) the enhanced ERP signal using the wPCA approach

Fig. 3 shows the time-frequency distributions for the above sample VEP and its wavelet-based VEP estimate, respectively. Fig. 3(a) corresponds to the original signal. Fig.3 (b) corresponds to the reconstructed single-trial signal by the combined method. The unpleasant stimuli appeared at 0 s. The same axis range for the amplitude is used here. Visual-related activity was clearly noticeable in the time-frequency distribution of the wavelet-based VEP estimate, whereas such activity could hardly be seen from the raw signal. Therefore, we concluded that the wavelet-based combined method could recover the evoked potential.

The mean VEPs of the 15 single trials under the three types of stimuli for the above subject are obtained for the PO8 site. Mean voltages in this region are assessed in the P300 (300-500 ms) and in the slow wave window (550-900 ms) [17-19]. Fig. 4 (a) shows the mean VEPs of the original single-trial responses at the PO8 site. Fig. 4(b)

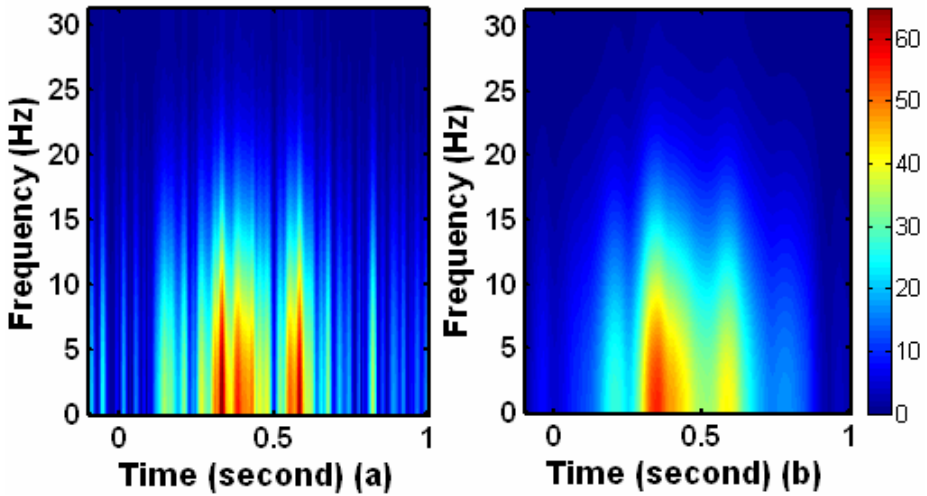


Fig. 3. Sample results for the time-frequency plot of a single trial of VEP: (a) corresponding to the original signal; (b) corresponding to the reconstructed single-trial signal by the combined method

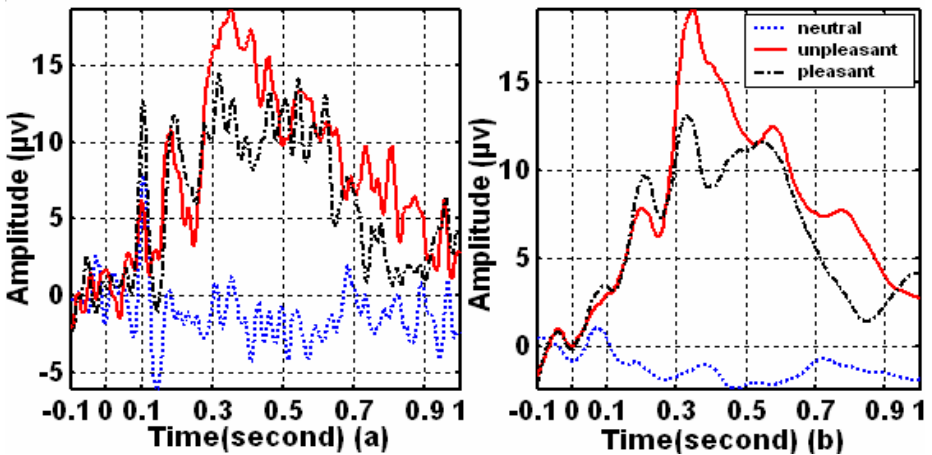


Fig. 4. Average VEPs at the electrode PO8 in response to three types of emotional stimuli for a subject: (a) Original single-trial responses; (b) the estimated VEPs by using the wPCA approach

shows the average VEP estimates obtained by the wPCA method. We can see that the VEP was composed of five components: a N100, a P200, a N200, a P300 component and a late positive slow wave. Here, we focus on the P300 and slow wave time window which indicates the sustained and high-level processing of salient visual stimuli [18].

4 Discussion

In this paper, a new wPCA approach is compared to WD, PCA and classical ensemble averaging using simulated data and actual recordings of emotion VEPs. Our objective is to investigate the effectiveness of these methods in estimating the overall morphology and, in particular, the P300 and the slow wave components in the experimental single-trial emotion VEPs. The results show that the VEPs obtained by the wPCA method could be used as a reliable, sensitive, and high-resolution indicator for clinical cognitive studies after only 15 trials of ensemble averaging.

With simulated data, we compare the wPCA method and the other two methods and demonstrate that the former has higher SNR and lower RMSE than the latter. The wPCA method has much more smooth estimates than the other methods. In addition, wavelet-based approaches are immune to the effect of trial-by-trial evoked potential variations.

When actual data are used, the wPCA method also improves the visualization of the single-trial ERPs compared with the other two methods. The results show greater P300 and slow wave amplitudes for unpleasant and pleasant pictures compared to neutral stimuli, indicating that motivationally relevant stimuli automatically directed attention resources, are processed more deeply and thus provoked an arousal-related enhancement of VEPs, which further supports the view that emotional stimuli are processed more intensely [16,18,19].

Characteristics ERPs can be captured by means of wavelet-based analysis, which can be further used for the detection and recognition of abnormalities in the human brain. ERP extraction results can be improved by the proposed wavelet-based combination method in this paper.

Acknowledgments. This work was supported by the open project of the State Key Laboratory of Cognitive Neuroscience and Learning and the open project of the Beijing Key Lab of Applied Experimental Psychology at the Beijing Normal University. The authors would like to thank Professor Senqi Hu in the Department of Psychology at the Humboldt State University, Arcata California for useful discussions.

References

1. Bradley, A.P., Wilson, W.J.: On wavelet analysis of auditory evoked potentials. *Clin. Neurophysiol.* 115, 1114–1128 (2004)
2. Iyer, D., Zouridakis, G.: Single-trial evoked potential estimation: Comparison between independent component analysis and wavelet denoising. *Clin. Neurophysiol.* 118, 495–504 (2007)
3. Kook, H., Gupta, L., Kota, S., Molfese, D., Lyytinen, H.: An offline/real-time artifact rejection strategy to improve the classification of multi-channel evoked potentials. *Pattern Recognition* 41, 1985–1996 (2008)
4. Dien, J., Beal, D.J., Berg, P.: Optimizing principal components analysis of event-related potentials matrix type, factor loading weighting, extraction, and rotations. *Clin. Neurophysiol.* 116, 1808–1825 (2005)
5. Quian-Quiroga, R., Garcia, H.: Single-trial event-related potentials with wavelet denoising. *Clin. Neurophysiol.* 114, 376–390 (2003)

6. Demiralp, T., Ademoglu, A., Istefanopoulos, Y., Basar-Eroglu, C., Basar, E.: Wavelet analysis of oddball P300. *Int. J. Psychophysiology* 39, 221–227 (2001)
7. Demiralp, T., Ademoglu, A., Schurmann, M., Basar-Eroglu, C., Basar, E.: Detection of P300 waves in single trials by the Wavelet transform (WT). *Brain and language* 66, 108–128 (1999)
8. Kobayashi, T., Kuriki, S.: Principal Component Elimination Method for the Improvement of S/N in Evoked Neuromagnetic Field Measurements. *IEEE Trans. Biomed. Eng.* 46, 951–958 (1999)
9. Thireou, T., Strauss, L.G., Dimitrakopoulou-Strauss, A., Kontaxakis, G., Pavlopoulos, S., Santos, A.: Performance evaluation of principal component analysis in dynamic FDG-PET studies of recurrent colorectal cancer. *Comput. Med. Imaging Graph* 27, 43–51 (2003)
10. Palaniappan, R., Ravi, K.V.R.: Improving visual evoked potential feature classification for person recognition using PCA and normalization. *Pattern Recognition Letters* 27, 726–733 (2006)
11. Polikar, R., Topalis, A., Green, D., Kounios, J., Clark, C.M.: Comparative Multiresolution Wavelet Analysis of ERP Spectral Bands Using an Ensemble of Classifiers Approach for Early Diagnosis of Alzheimer's Disease. *Comput. Biol. Med.* 37, 542–556 (2007)
12. Jolliffe, I.T.: *Principal Component Analysis*, 2nd edn. Springer, New York (2002)
13. Turner, S., Picton, P., Campbell, J.: Extraction of short-latency evoked potentials using a combination of wavelets and evolutionary algorithms. *Med. Eng. Phys.* 25, 407–412 (2003)
14. Masahiko, N.: Waveform Estimation from Noisy Signals with Variable Signal Delay Using Bispectrum Averaging. *IEEE Trans. Biomed. Eng.* 40, 118–127 (1993)
15. Yu, X.H., He, Z.Y., Zhang, Y.S.: Time-varying adaptive filters for evoked potential estimation. *IEEE Trans. Biomed. Eng.* 41, 1062–1071 (1994)
16. Zou, L., Zhou, R.L., Hu, S.Q., Zhang, J., Li, Y.S.: Single Trial Evoked Potentials Study during an Emotional Processing Based on Wavelet Transform. In: Sun, F., Zhang, J., Tan, Y., Cao, J., Yu, W. (eds.) *ISNN 2008, Part I. LNCS*, vol. 5263, pp. 1–10. Springer, Heidelberg (2008)
17. Keil, A., Müller, M.M., Gruber, T., Stolarova, M., Wienbruch, C., Elbert, T.: Effects of emotional arousal in the cerebral hemispheres: a study of oscillatory brain activity and event-related potentials. *Clin. Neurophysiol.* 112, 2057–2068 (2001)
18. Cuthberg, B., Schupp, H., Bradley, M., Birbaumer, N., Lang, P.: Brain Potentials in Affective Picture Processing: Covariation with Autonomic Arousal and Affective Report. *Biol. Psychol.* 52, 95–111 (2000)
19. Herbert, B.M., Pollatos, O., Schandre, R.: Interoceptive Sensitivity and Emotion Processing: An EEG study. *Int. J. Psychophysiology* 65, 214–227 (2007)

Fourier Volume Rendering on GPGPU

Degui Xiao, Yi Liu, Lei Yang, Zhiyong Li, and Kenli Li

School of Computer and Communication, Hunan University,
Changsha 410082, China

Abstract. Fourier Volume Rendering (FVR) is a volume rendering technique with lower computational complexity of $O(N^2 \log N)$ for an N^3 data array. A new FVR algorithm is proposed through expanding Fourier Projection-Slice Theorem into High-Dimension and mapping the pipeline totally on GPU. A windowed-sinc function is used as reconstruction filter to implement higher-order interpolation and reduction of samples is executed on GPU in parallel, which meets the architecture of Heterogeneous multi-core. The rendering is accelerated by a factor of 7 when rendering image's resolution is larger than 512×512 .

Keywords: Volume rendering, Fourier transform, Higher-order interpolation, GPGPU.

1 Introduction

GPGPU means General Purpose Graphic Process Unit, which focuses on the floating-point operation and parallel processing of GPU to deal with the general computation rather than three-dimension graphics applications. To speed up the volume rendering with GPGPU, especially for large-scale volume data scientific visualization, has far-reaching meaning.

At present, volume rendering can be divided into two categories:

The method based on screen space or image space: a ray is cast for each pixel on the screen, with uniform sampling and composition of the volumetric data along the ray, such as Ray-casting [1] and Shear-Warp [2].

The method based on object space: the volume is traversed either back-to-front or front-to-back, blending each scalar into the projection plane, e.g. three-dimension texture mapping [3] and Splatting [4].

Algorithms in both categories operate in the spatial domain and have to travel every sample in the data set with the complexity of $O(N^3)$. Although some adaptive techniques can avoid accessing every sample of the data set, for example, in Ray-casting one can terminate rays when the ray opacity values are close to unity, or hierarchical data structures can be used to avoid visiting volumes of empty space, they relies greatly on the structure of the data set. Different from the methods above, Fourier Volume Rendering, firstly proposed by [5], operates in frequency domain to compute project slices of three-dimension discrete data and reduces the complexity to $O(N^2 \log N)$.

We present mapping FVR algorithm to GPGPU to significantly accelerate the rendering performance. An overall pipeline of hardware accelerated frequency domain volume rendering is presented. The paper is organized as follows. Section 2 is a simple

introduction of relative works. Section 3 introduces Fourier Slice Theorem and expands it into high-dimension. Section 4 shows the implementation of this algorithm on GPGPU in detail and the experimental results. Conclusion is in section 5.

2 Related Works

Fourier volume rendering has been further improved in both rendering speed and image quality since it's first presented.

[6] extended this work with depth cues and shading performing calculations in the frequency domain during slice extraction. Illumination models for FVR were studied in the work of [7]. They describe methods to integrate diffusion lighting into FVR. One approach is based on gamma corrected hemispherical shading and is suitable for interactive rendering of fix light sources. Another technique uses spherical harmonic functions and allows lighting using varying light sources. [8] proposed frequency volume rendering based on wavelet transform. The Fast Hartley Transform (FHT), proposed by [9] as an alternative to FFT, produces real output for a real input, and is its own inverse. Therefore for FVR, the FHT is more efficient in terms of memory consumption.

All the algorithms above are implemented on CPU, compared to the majority of the volume rendering in spatial domain which have used one of the current GPU features, programmability. As a result, even the computational complexity of traditional method is higher than the FVR, the rendering speed is much faster. The mathematical structure of FVR algorithm, especially the Fast Fourier Transform prevents its use on the GPU. The core of Fast Fourier Transform is butterfly computation, in order to implement it on GPU, [10] proposed "Stream-FFT". Before each butterfly computation, an internal exchange should be executed during the input data, therefore the two values which do the butterfly computation would be adjacent, to meet the requirements of GPU computing. [11] used the four channels of the texture to store the filter, to execute the high order interpolation.

3 Method

Volume rendering can be seen loosely as the inverse process of tomographic reconstruction. In tomographic reconstruction, the goal is to compute the unknown distribution of the three-dimension data set from the projections which are generated by the scan of the X-ray in different angles. In contrast, the volume rendering is to generate the projection in any desired angle with the given distribution of the volume data.

One way to achieve this reconstruction is through the use of the Fourier Projection-Slice Theorem, which means the 1D Fourier transform of a projection of an object at some angle is a slice of the 2D Fourier Transform of the object at the same angle. Graphically, this is illustrated in Fig. 1.

Expanding the theorem into High-Dimension, we get a new theorem as follow:

Theorem 1. After transform the N dimensional function $f(x_1, x_2, \dots, x_N)$ into frequency domain with Fourier Transform, we get $F(U_1, U_2, \dots, U_N)$. If we fix one dimension U_k in F , then get an $N-1$ dimension function F' through the origin point. After an inverse Fourier Transform, a $N-1$ dimension function f' in spatial domain can be generated from F' . So f' should be the integral of $f(x_1, x_2, \dots, x_N)$ along the x_k direction.

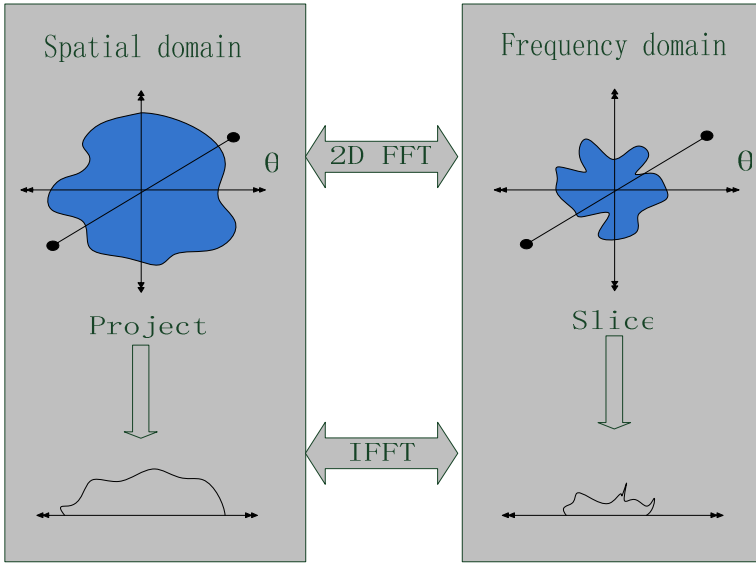


Fig. 1. The Fourier projection-slice

Proof. Suppose $g(x_1, x_2, \dots, x_N)$ to be a N dimensional function, the projection of g along x_N direction can be illustrated as integral:

$$f(x_1, x_2, \dots, x_{N-1}) = \int_{-\infty}^{\infty} g(x_1, x_2, \dots, x_N) dx_N \tag{1}$$

$N-1$ dimensional Fourier Transform is defined as follow :

$$F(U_1, U_2, \dots, U_{N-1}) = \int \dots \int_{-\infty}^{\infty} f(x_1 \dots x_{N-1}) e^{2\pi i(U_1 \times x_1 + \dots + U_{N-1} \times x_{N-1})} dx_1 \dots dx_{N-1} \tag{2}$$

We can get (3) from (1) and (2):

$$F(U_1, U_2, \dots, U_{N-1}) = \int \dots \int_{-\infty}^{\infty} g(x_1 \dots x_N) e^{2\pi i(U_1 \times x_1 + \dots + U_{N-1} \times x_{N-1})} dx_1 \dots dx_{N-1} dx_N \tag{3}$$

Then we add one dimension x_N , but x_N should be fixed, which means x_N equals to 0:

$$F(U_1, U_2, \dots, U_{N-1}) = \int \dots \int_{-\infty}^{\infty} g(x_1 \dots x_N) e^{2\pi i(U_1 \times x_1 + \dots + U_N \times x_N)} dx_1 \dots dx_{N-1} dx_N \Big|_{x_N=0} \tag{4}$$

And the N dimension Fourier Transform is:

$$F(U_1, U_2, \dots, U_{N-1}, U_N) = \int \dots \int_{-\infty}^{\infty} f(x_1 \dots x_N) e^{2\pi i(U_1 \times x_1 + \dots + U_N \times x_N)} dx_1 \dots dx_N \tag{5}$$

So (4) is a representation of N dimensional Fourier Transform:

$$F(U_1 \dots U_N) = FT_N \{ f(x_1 \dots x_N) \} = G(U_1 \dots U_{N-1}, 0) \tag{6}$$

At last, the theorem has been proved:

$$f(x_1 \dots x_N) = IFT_{N-1} \{ G(U_1 \dots U_{N-1}, 0) \} \tag{7}$$

Where FT_N is N dimensional Fourier Transform, IFT_{N-1} is $N-1$ dimensional inverse Fourier Transform.

4 The Implementation on GPGPU

Because of the hardware limitations, previous algorithm achieved by Fourier Volume Rendering is implemented totally on CPU, or partly on GPU. With the development of modern GPU, the floating-point operations capability and bandwidth of GPU has been far beyond the CPU, also it provides three-dimension texture and double precision floating-point which build the foundation of FVR on GPGPU.

Based on NVIDIA CUDA architecture, we propose a new Fourier Volume Rendering algorithm: only the operation of reading data and some simple condition judgment are handle by CPU, all the computation for the large amounts of data is totally finished on GPU, which meets the architecture of Heterogeneous multi-core.

CUDA (Compute Unified Device Architecture) is a new computing architecture proposed by NVIDIA, which provide a new computing ability on data-intensive application with the powerful ability of the GPU. Through the standard C language, CUDA provides a large number of high-performance as well as concise instructions of the development process, thus allowing developers to create a solution that would consume less time for data-intensive processing to provide a precise enough result. In CUDA, GPU is viewed as a computing device which could execute many "Thread" in parallel. The Threads which share data and synchronize memory access are organized as "Block". Because the limitation of the Thread number in a Block, the Blocks which have the same dimension and can execute the same program are called "Grid" [12]. Based on the parallel execution of each Thread, this paper distributes the operations between the samples to every Thread, which takes full advantage of SIMD (Single Instruction Multiple Data) computation on GPGPU.

With several steps, Fourier Volume Rendering is implemented on GPU, only the first step we need to transfer the data from CPU to GPU, the left steps are all finished on GPU, which avoids the transferring samples between CPU and GPU. The main steps are:

1. Pre-process: samples are transformed from spatial domain to frequency domain.
2. High order interpolation: high order interpolation is executed on projection slice to avoid aliasing and ghosting artifacts.
3. Inverse Fourier Transform: samples after resample are transformed back to spatial domain
4. Rendering: slice is normalized and rendered to frame buffer.

And the algorithm flow is illustrated in Fig. 2.

4.1 Pre-process

With three-dimension Fourier Transform, discrete samples are transformed from spatial domain into frequency domain. As previous GPU lacks of support on three-dimension texture, this step is usually completed on CPU. With the development of GPU, three-dimension is supported by the latest GPU of NVIDIA. So during the pre-process step, the samples in memory are transferred into "global memory" which can be accessed by CUDA, then with FFT library in CUDA, a three-dimension Fast Fourier Transform is executed on GPU in parallel. Some results of execution time on GPU compared with that on CPU with FFTW are shown in table 1.

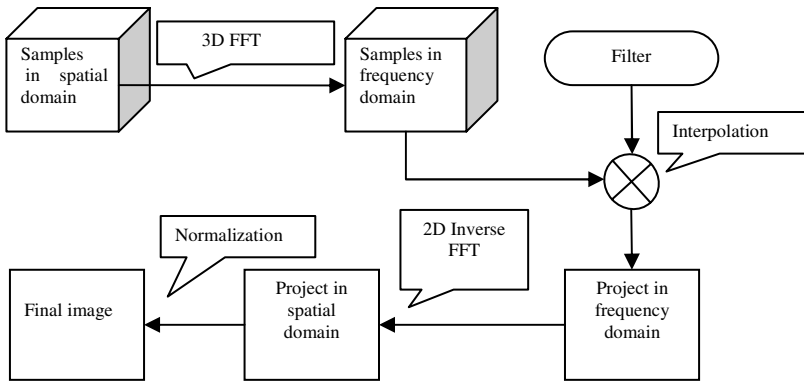


Fig. 2. Flow of FVR on GPGPU

Table 1. Fourier Transfer between CPU and GPGPU

Data size	Time(ms)	
	FFTW	GPGPU
32768 (32 ³)	0.66	3.86
262144 (64 ³)	9.66	5.33
2097152 (128 ³)	200.68	23.95
16777216 (256 ³)	1122.49	114.88

FFTW is developed by M.I.T computer science to compute Fast Fourier Transform with different dimension. When the data size is small, the time consumed on GPGPU is a litter more than CPU, but with the increase of data size, GPGPU can attain nearly 10 times speedup.

4.2 High-Order Interpolation

The samples after pre-process step should be interpolated on the projection slice which is oriented perpendicular to the viewing direction. This step is finished by a reconstruction filter in frequency domain. Resample in frequency domain has to follow Nyquist Theorem. Here we use windowed-sinc function as our reconstruction filter and the interpolation is carried on by the convolution of filter and samples on GPU. Because the sinc function is an unacceptable reconstruction filter due to its infinite extent and needs to be replaced by some finite-extent approximation. A better choice for a clipping function in the frequency domain is the Blackman-Window. With Cut-off frequency f and the length of the filter M , we get the expression of the filter as follow:

$$K(i) = \frac{\sin(2\pi f(i - M/2))}{i - M/2} [0.42 - 0.5 \cos(2\pi i / M) + 0.8 \cos(4\pi i / M)]$$

According to Convolution Theorem, convolution in frequency domain equals to product in spatial domain. So the convolution between samples and filter in frequency

domain can be obtained by the product in spatial domain. Samples and the filter are transformed via Inverse Fourier Transform then they are loaded by Block. Each product operation is handled by one Thread in a Block, and after producing product of each point, the result is sent back to frequency domain. The interpolation step is finished after Fourier Transform which transforms the result back to frequency domain.

The code fragment of high order interpolation is as follow:

```
high_order_interpolation(indata ,M, data_size,outdata)
begin
  for i=0 to M do
    if( i-M/2 == 0 ) then
      begin
        kernel[i].x=2*PI*f;
      end
    else
      begin
        kernel[i].x=sin(2*PI*f(i-M/2))/(i-M/2)
        kernel[i].x=kernel[i].x*(0.42-0.5*cos(2*PI*i/M)+0.
          8*cos(4*PI*i/M))
        kernel[i].y=0;
      end;
    endif;
    for i=0 to data_size do
      begin
        out-
        data[i]=kernel[i].x*indata[i].x-kernel[i].y*ind
          ata[i].y,kernel[i].x*indata[i].x+kernel[i].y*kerne
            l[i].y
          end;
        CUFFT_SAFE_CALL(cufftExecC2C(FFTplan, (cufftComplex*
          )outdata, (cufftComplex*)outdata,CUFFT_INVERSE) );
      end;
    end;
  end;
end;
```

4.3 Inverse Fourier Transform

After second step, the slice in frequency domain is generated. In order to transform the slice back to spatial domain, an Inverse Fourier Transform in CUDA FFT library is executed. To generate an image with size $N \times N$, the input of this step is an array of N^2 , the computing complexity of the transform is $O(N^2 \log N)$. During the pre-process step, the Fourier Transform has to deal with the three-dimension data with the computing complexity of $O(N^3 \log N)$. But for each data set the pre-process step only needs to be executed once, all the following steps only deal with the two-dimension data. That means the interpolation and inverse transform step work on one slice of the volume data, which ensures the computing complexity of the algorithm to be $O(N^2 \log N)$.

From the theorem which has been proved above, the slice after inverse Fourier Transform equals to the project of the original volume data in spatial domain.

4.4 Rendering

The last step of this algorithm is going to render the image. All the samples in a slice should be normalized before rendering. The key of normalization is to find the maximum value and minimal value of the samples, what we have done is to get the values through a reduction operation which looks like an inverted tree, as shown in Fig. 3.

In the tree-based approach, we use multiple thread blocks, each thread block reduces a portion of the array. The problem is how to synchronize across all thread blocks when CUDA provides no global synchronization. Our solution is to decompose reduction operation into multiple kernels, so the computation is decomposed into different levels and codes in every level are totally same. And the kernel is invoked recursively until we get the extremum. Then the samples are normalized between [0,255].

Finally, the samples are assigned corresponding gray-scale values. Images are directly rendered by sending the results to Pixel Buffer Object that registered before, which avoids transferring data back to CPU.

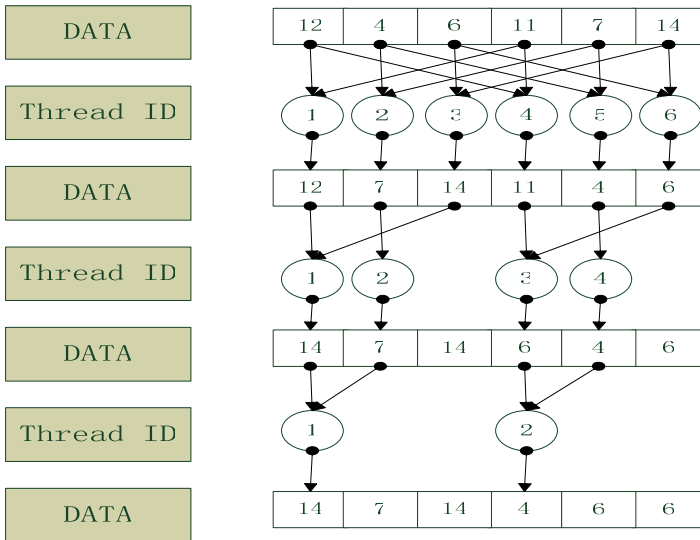


Fig. 3. The operation to get extremum on GPGPU

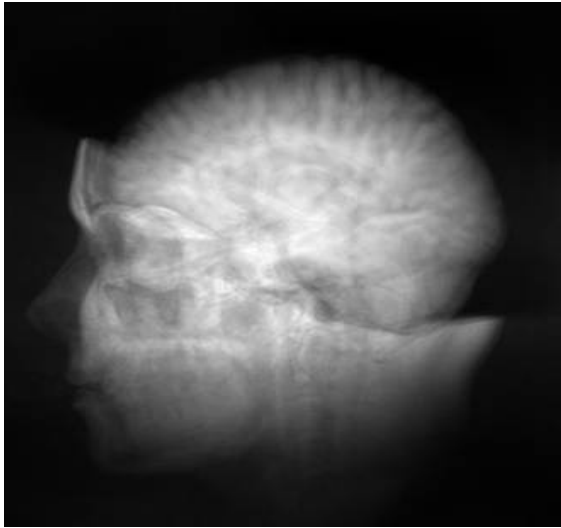
4.5 Results

The algorithm is running on Intel Core2Duo with NVIDIA Geforce 8800GTS, the data set we use is a medical data set of size 2563. The performance of FVR relies on the size of the project slice, so the images are rendered in different resolutions: 256×256, 512×512 and 1024×1024. Some results are listed in Table 2 and shown in Fig. 4.

Table 2. Experimental results

	CPU			GPGPU		
	256 ²	512 ²	1024 ²	256 ²	512 ²	1024 ²
Resolution	256 ²	512 ²	1024 ²	256 ²	512 ²	1024 ²
Interpolation (ms)	9.58	29.71	104.35	7.26	15.24	24.26
IFFT (ms)	4.76	19.99	93.82	0.43	0.88	2.12
Normalization (ms)	2.12	5.33	22.26	0.79	1.60	4.79
Total (ms)	16.46	55.04	220.43	8.48	17.72	31.17

From the results we can see that the interpolation consumes most of the rendering time, but with the algorithm mapped on the GPGPU, a speed-up factor of approximately 7 is achieved.

**Fig. 4.** Medical data set

5 Conclusion

With the programmability and parallelism of modern graphic hardware, a new Fourier Volume Rendering based on GPGPU is proposed to accelerate the rendering speed of three-dimension volume data. The samples are transformed into frequency domain during the pre-process and store in GPU as three-dimension texture. Adopting the parallel computing ability of GPU, high order interpolation is finished rapidly in frequency domain, and further enhancement is gained by normalization with thread blocks. The performance of Fourier Volume Rendering is not dependent on the size of the data set but the size of the slice resolution, which can make it be widely used for large volume data visualization.

This algorithm only works on NVIDIA CUDA architecture, so NVIDIA graphics hardware is necessary. Future work includes mapping to other platforms, e.g. some high level shading implementation using shader language. Also quality of the resulting images should be improved by integrating lighting.

Acknowledgements

This work is supported by the National Natural Science Foundation of China under Grant Nos. 90715029.

References

1. Levoy, M.: Display of Surfaces from Volume Data. *J. IEEE Comp. Graph. & Appl.* 8, 29–37 (1988)
2. Lacroute, P., Levoy, M.: Fast Volume Rendering Using a Shear-Warp Factorization of the Viewing Transformation. In: *SIGGRAPH 1994*, Orlando, pp. 451–458 (1994)
3. Cabral, B., Cam, N., Foran, J.: Accelerated Volume Rendering and Tomographic Reconstruction Using Texture Mapping Hardware. In: *Symposium on Volume Visualization 1994*, Tysons Corner, pp. 91–98 (1994)
4. Westover, L.: Foot Print Evaluation for Volume Rendering. In: *17th Annual Conference on Computer Graphics and Interactive Techniques*, Dallas, pp. 367–376 (1990)
5. Malzbender, T.: Fourier Volume Rendering. *J. ACM Transactions on Graphics* 12, 233–250 (1993)
6. Totsuka, T., Levoy, M.: Frequency Domain Volume Rendering. In: *20th Annual Conference on Computer Graphics and Interactive Techniques*, Anaheim, pp. 271–278 (1993)
7. Entezari, A., Scoggins, R., Möller, T., Machiraju, R.: Shading for Fourier Volume Rendering. In: *Proceedings of Symposium on Volume Visualization and Graphics 2002*, Boston, pp. 131–138 (2002)
8. Westenberg, M.A., Roerdink, J.B.T.M.: Frequency Domain Volume Rendering by the Wavelet X-Ray Transform. *J. IEEE Transactions on Image* 9, 1249–1261 (2002)
9. Bracewell, R.N., Buneman, O., Hao, H., Villasenor, J.: Fast Two-Dimensional Hartley Transform. *Proceedings of the IEEE* 74, 1282–1283 (1986)
10. Jansen, T., Rymon-Lipinski, B.V., Hanssen, N., Keeve, E.: Fourier Volume Rendering on the GPU Using a Split-Stream-FFT. In: *Proc. VMV 2004*, Stanford, pp. 395–403 (2004)
11. Viola, I., Kanitsar, A., Groller, M.E.: GPU-based Frequency Domain Volume Rendering. In: *Proceedings of SCCG 2004*, Budmerice, pp. 49–58 (2004)
12. NVIDIA. *CUDA Programming Guide 2.0*, http://www.nvidia.com/object/cuda_get

An Improved Population Migration Algorithm for the Prediction of Protein Folding

Huafeng Chen and Jianyong Wang*

College of Science, Huazhong Agriculture University, Wuhan 430070, China
chenhf123@mail.hzau.edu.cn

Abstract. This paper discusses the calculating problem about protein-folding lattice model and bings up the Improved Population Migration Algorithm. The algorithm adds the idea of the Genetic Algorithm and the Immune Algorithm into the framework of the Population Migration Algorithm. The experiment results show that the Improved Population Migration Algorithm has strong global search capability and stability and can be obtained better solution than the existing algorithm.

Keywords: Protein-folding lattice model, Improved population migration algorithm, Stability.

1 Introduction

The problem of protein folding is also the problem of protein structure prediction. It is one of the core issues in the field of bioinformatics and the center issues of molecular biology , it is also an important task to solve the problem in after the era of gene protein engineering. In the early 1960s, Anfinsen introduce the famous thesis: natural protein conformation is the lowest energy conformation. Since then, it is a reasonable assumption of thermodynamics to use of energy minimization method to predict the protein structure.

At present, the problem of protein structure prediction that is still a problem to be resolved, the main difficulty is the folding space growth in exponential as the length of protein sequence growth. People do effort form two sapects to slove the problem: (1) Physical models and mathematical models are simplified in the conditon of maintaining the accuracy of the conditions, such as lattice model [1], off-lattice model [2], etc. (2) It look for predicting the protein structure of global optimization methods, for example Immune Algorithm(IA), Genetic Algorithm(GA) , Simulated Annealing (SAand so on [3] . This article makes some improvement about the Population Migration Algorithm and apply the Improved Population Migration Algorithm in off-lattice model.

The Population Migration Algorithm(referred to PMA) [4][5] introduced by Yonghua Zhou and Zongyuan Mao the scholar of china is global optimization algorithm by simulating population migration mechanism.The algorithm mainly

* Corresponding author.

simulates the mechanism which the population shift follow the economic center and increasing pressure of population, the former make the algorithm to choose a better search area and the latter can be avoid the algorithm into a local optimization, therefore, the process of search show the features alternating centralized and distributed search.

2 Off-Lattice Model of Protein Folding

20 varieties of amino acids are divided into two types according to the hydrophilicity or hydrophobicity: A(hydrophobicity) and B (hydrophilicity), so using A and B to represent protein sequence. The two adjacent amino acid are linked with a unit bond. The protein sequence $P = p_1p_2 \cdots p_n$ represent length of $n - 1$ and arbitrary folding line, so the sequence form $n - 2$ angles between bond and bond. The distance is about the function of the bond angle, as the bond length is the length of unit. The function of distance r_{ij} is:

$$r_{ij} = ((1 + \sum_{k=i+1}^{j-1} \cos(\sum_{l=i+1}^k \theta_l))^2 + (\sum_{k=i+1}^{j-i} \sin(\sum_{l=i+1}^k \theta_l))^2)^{1/2} \tag{1}$$

Energy E of a sequence of length n is defined as follow :

$$E = \sum_{i=2}^{n-1} V_1(\theta_i) + \sum_{i=1}^{n-2} (\sum_{j=i+2}^n V_2(r_{ij}, \xi_i, \xi_j)) \tag{2}$$

$$V_1(\theta_i) = \frac{1}{4}(1 - \cos\theta_i), V_2(r_{ij}, \xi_i, \xi_j) = 4(r_{ij}^{-12} - C(\xi_i, \xi_j)r_{ij}^{-6}),$$

For

$$C(\xi_i, \xi_j) = \frac{1}{8}(1 + \xi_i + \xi_j + 5\xi_i\xi_j), \xi_i = \begin{cases} 1 & , p_i = a; \\ -1 & , p_i = b. \end{cases} (i = 1, 2, \dots, n) \tag{3}$$

Forming equation (3), we can know that for the time correlation coefficient $C(\xi_i, \xi_j)$ is 1 when non-adjacent residues are AA , when AB for the correlation coefficient $C(\xi_i, \xi_j)$ is $-1/2$ when non-adjacent residues are AB and the correlation coefficient $C(\xi_i, \xi_j)$ is $1/2$ when non-adjacent residues are BB, therefore , two hydrophobic residues have a very strong gravity , the two hydrophilic residues have slight gravity , between hydrophobic residue and hydrophilic residue has a slight repulsion , this reflects the true property of the protein in some extent.

3 Application of the Population Migration Algorithm in Protein Off-Lattice Model of Folding

3.1 The Idea of the Population Migration Algorithm

The framework of the Population Migration Algorithm as follow:

- 1) People migrate in local area;

- 2) The preferential region attract people immigrate to there;
- 3) People immigrate into preferential region until the population pressure reach a certain limit;
- 4) People move out form preferential region and look for new opportunities.

In this process, on the one hand people get together in preferential region, on the other hand people move away form preferential region as the population pressure. The migration of the population is the process which people look for preferential region in constant accumulation and proliferation of contradiction.

3.2 Application of the Population Migration Algorithm in Protein Off-Lattice Model of Folding

According to compare with the migration algorithm results (local optimal value), we found their configuration of corresponding are very similiary,we do 9 times continual calculation using the sequence of BBABBABB ,Although the energy values are not equal in the 9 times iteration, each bond angle of BAB always in a certain range. We belive the part configuration that have roughly equal vaule except the direction (sign) have a good property , for exemple each configuration of the bond angle 3 and 6. Therefore , using the idea of immune algorithm and introduction of antibody memory mechanism, m local optimal points that are prdouced m times iteration of population migration algorithm as initial structure population , the small part of the configuration (the part of good characters) are retained to conduct genetic opereation, so the convergence rate of solution is faster and more stable.

3.3 A Mixed Population Migration Algorithm (Immune-Genetic-Migration Algorithm)

Step 1 : In the search space prdouce N points of uniform random (protein configuration),let the i -th regional center $center^i = x^i$, for each i ; identify the upper and lower bounds ($center^i \pm \delta^i$) of the i th regional ,for $i = 1, 2, \dots, N$; computing energy E of every points, and initializing the best record value and the best record point.

Step 2 : The population migrate in each regional; moving each points in uniform random: $x^i = 2\delta \cdot rand(*) + (center^i - \delta)$, $rand(*)$ is the random function. If $x_j^i > b_j$, let $x_j^i = 2\pi - x_j^i$; If $x_j^i < a_j$, then let $x_j^i = 2\pi + x_j^i$; If $x_j^i = b_j$ or $x_j^i = a_j$, then give up it. Computing energy E for each points, recording the best value and point.

Step 3 : If The number of population movement l is less than the number of pre-specified then switch to step 2.

Step 4 : The population migration : It is the lowest energy point (That is the best record point) as central to identify the preference region. There are N points be produced in random and uniform to replace the original points in this region, then computing energy E of each point, and recording the best value and point.

Step 5 : Shrinking the preference region: $\delta = (1 - \Delta)\delta$ (Δ : shrinkage coefficient, $0 < \Delta < 1$), population migrate in preference region: It is the lowest energy

point(That is the best record point) as central to identify the preference region, and there are N points be produced in random and uniform to replace the original points in this region, then computing energy E of each point, and recording the best value and point.

Step 6 : If $\delta > \alpha$ (α is the parameter of population pressure, and it is a positive small number pre-given), then swith to step 5.

Step 7 : Reporting result.

Step 8 : Population extend: Producing N points x^1, x^2, \dots, x^N to replace original points in search region.

Step 9 : Times of iteration is $m + 1$,if times of iteration less than the number of pre-designated, then switch to step 2.

Step 10: The configuration memory bank that are produced by m times iteration make the initial structure of the population.

Step 11: Adjustment of adaptive value :

$$f(k) = (\alpha + 0.5)^{\text{sign}(c-\gamma)\text{sign}(\frac{f(k)}{f_{max}}-\alpha)} f(k)$$

$$\text{sign}(x) = \begin{cases} 0 & , x < 0; \\ 1 & , x \geq 0. \end{cases}$$

f_{max} is the minimum adaptive value in population, and α is the threshold value of ratio which adaptive value of better antibody defined divides by f_{max}, γ is the concentration of threshold value of better antibody, c is the total concentration of better antibody, f_k is the adaptive value of $NO.k$ antibody.

Step 12: Genetic operation.

(1) select : compared with the adaptive value(ene. rgy) of initial antibody ueeing sorting selection, and the smaller $m/20$ individuals directly into next generation.

(2) cross : P_1 and P_2 are randomly selected from population as parent, and according to the probability $P_c(= 0.5)$ to cross operation (Randomly select 2 individuals, and comparing same part of the energy, then selecting a better random to generate cross i ; P_1 and P_2 are divided into two parts form cross i , and then changing the correspond part.)

(3) variation: Selecting one individual form population in random,and then carrying out mutation in the probability $P_m(= 0.03)$.(Selecting one individual P_1 in random , then generating variant i in random; if variant i of P_1 is 1 replacing it with 0 ,and if it is 0 replacing with 1.)

Step 13: Repeating step 11 and 12 until the number of cycle $M > 20$ of the best individual not change.

Step 14: Producing the size of a $m/10$ antibody memory bank,selecting the best of $m/20$ antibody into the antibody bank and replacing the worst of $m/20$ antibody. The antibody in antibody bank is a part of initial population and others produced in random.

Step 15: Repeat steps 11-14 until Antibody memory bank to update the number of $L > 5$.

4 Numerical Experiment Results

We use the software of VC++6.0 to achive the Population Migration Algorithm (PMA) and the Improved Population Migration Algorithm (IPMA),and numerical experiments are carrie out in the microcomputer of the pentium (R) 4 3.06GHz and 1 G memory, we chose the following sequence of the protein as example, and comparing with the results of Genetic Algorithm and Simulated Annealing[3], and showing results.

- ♣₁(12) BAABBBBBABAB
- ♣₂(13) BABBABBABBABB
- ♣₃(15) BBABBAABBBBAABB
- ♣₄(17) BBBAABBAABBBBBBBB
- ♣₅(18) BBABBABABABBBBABB
- ♣₆(20) BBABBABABABABBABBABB

Energy E is the minimum energy in 10 serial calculation,

Algorithm	GA	SA	PMA	IPMA
Sequence	E_{min}	E_{min}	E_{min}	E_{min}
♣ ₁	- 5.91804	- 6.21719	- 6.3616	- 6.6342
♣ ₂	- 8.08493	- 6.72458	- 7.26548	- 8.1733
♣ ₃	- 16.3377	- 15.8845	- 16.6732	- 17.3329
♣ ₄	- 15.3119	- 14.9615	- 15.3120	- 15.4988
♣ ₅	- 17.1135	- 18.6404	- 18.0993	- 18.8488
♣ ₆	- 13.6780	- 17.5107	- 17.6233	- 18.5177

The results of experiment are more satisfactory. The PMA is better than GA and SA for the sequence ♣₁♣₃♣₄♣₆; The result of sequence ♣₂ is worse than GA , but it is better than SA. The result of sequence ♣₅ is worse than SA , but it is better than GA . The results of IPMA is significantly better and more stable than GA and SA.We belive the PMA and IPMA are applied to predict the protein space structure.

5 parameters are set in PMA and IPMA, N : Population size, l : time of population migration , Δ : shrinkage factor, a : population pressure parameters , m : time of iterations. Numerical experiments show that the result is better if the population size N set larger, usually seting $N = 1000$. It can search much fully and increase probability of finding global optimum , If we increase l and m and reduce Δ and a , and at the same time good Precision is prduced by reducing Δ and a .

5 Conclusions

The IPMA,a probability-based search algorithm, well slove the phenomenon of degradation that exit in the existing algorithm, and the convergence rate has improved significantly. At the same time, the algorithm have better global search capability and stability than Genetic Algorithm and Simulated Annealing in of optimization off-lattice model, and convergence rate and the stability of the solution are better than Polulation Migration Algorithm.

Acknowledgments. Bond of off-lattice model can rotate in every direction, so the model is similar to the real protein compared with the lattice model. Further research on PMA and IPMA can focus on the following:

- (1) Using the simplified model that is better reflect real characteristics of protein, for example, recently an improved off-lattice model is put forward by Chen Mao and Huang Wen-qi [6];
- (2) Extending PMA in IPMA Algorithm to solve problem of protein of three-dimensional lattice model and off-lattice model.

References

1. Lau, K., Dill, A.: A Lattice Statistical Mechanics Model of the Conformational and Sequence Spaces of Proteins. *Macromolecules* 22, 3986–3997 (1989)
2. Hinds, D.A., Levitt, M.: Exploring Conformational Space with a Simple Lattice Model for Protein Structure. *J. Mol. Biol.* 243, 668–682 (1994)
3. Niu, X.H., Li, N.N.: Immune Algorithm for the Prediction of Protein Folding. *J. of Math.* 24, 313–316 (2004)
4. Zhou, Y., Mao, Z.: A New Search Algorithm for Global Optimization: Population Migration Algorithm. *Journal of South China University of Technology* 31, 1–5 (2003)
5. Zhou, Y.Y., Mao, Z.Y.: A New Algorithm for Population Migration Optimization. *Journal of South China University of Technology* 31, 41–43 (2003)
6. Chen, M., Huang, W.Q.: Heuristic Algorithm for Off-lattice Protein Folding Problem. *J. Zhejiang Univ.* 7, 7–12 (2006)
7. Backofen, R.: The Protein Structure Prediction Problem: A Constraint Optimization Approach using a New Lower Bound. *Constraints* 6, 223–255 (2001)
8. Kim, S.Y., Lee, S.B., Lee, J.: Structure optimization by conformational space annealing in an off-lattice protein model. *Phys. Rev.* 72, 11–16 (2005)
9. Liang, F.M.: Annealing Contour Monte Carlo Algorithm for Structure Optimization in an Off-lattice Protein Model. *J. Chem. Phys.* 120, 6756–6763 (2004)
10. Hsu, H.P., Mehra, V., Nadler, W., Grassberger, P.: Growth-based Optimization Algorithm for Lattice Heteropolymers. *Phys. Rev. E* 68, 11–13 (2003)
11. Hsu, H.P., Mehra, V., Nadler, W., Grassberger, P.: Structure Optimization in an Off-lattice Protein Model. *Phys. Rev. E* 68, 77–80 (2003)
12. Ising, E.: Beitrag Zur Theorie des Ferromagnetismus. *Z. Physik.* 31, 253–258 (1925)
13. Heisenberg, W.: Zur Theorie des Ferromagnetismus. *Z. Physik.* 49, 619–636 (1928)

Gene Sorting in Differential Evolution

Remi Tassing¹, Desheng Wang¹, Yongli Yang¹, and Guangxi Zhu^{1,2}

¹ Department of Electronics and Information Engineering,

² Wuhan National Laboratory for Optoelectronics,

^{1,2} Huazhong University of Science and Technology, Wuhan 430074, China

tassingremi@gmail.com, dswang@hust.edu.cn,

yangyongli.wh@gmail.com, gxzhu@hust.edu.cn

Abstract. Gene sorting is a method proposed in this article that consists of ordering trial vector's component in differential evolution (DE). This method tends to significantly increase the convergence speed of DE with just a little modification on the original algorithm. A benchmark set of 18 functions is used for comparing both algorithms. Most importantly, the proposed methods can be incorporated in other variants of DE to further increase their respective speeds; Iterated Function System Based Adaptive Differential Evolution (IFDE) is used in this paper as a variant example and it is about 5 times faster for 30-dimension problems.

Keywords: Differential evolution, Global optimization, Convergence speed, Gene sorting.

1 Introduction

The purpose of global optimization is to find the point x^* minimizing a function $f(x)$, generally called objective or cost function, defined from a set $\Omega \subseteq \mathbb{R}^n$ to \mathbb{R} . A priori, there is no restriction on $f(\cdot)$, it can be continuous or not, differentiable or not, convex or not and with multiple local minima. In certain engineering fields, the function to be optimized might even be seen as a black box that does not have an explicit expression. Hence, classic optimization techniques such as linear programming (LP), the method of steepest-descent (SD) or least mean-squares (LMS) cannot be applied.

In contrast, genetic algorithms (GA) are a special group of optimization methods that depend less on the objective function properties. Hence, they have gain an increasing attention in the past few decades by their effectiveness and robustness in solving complex optimization problems in various research fields.

Differential Evolution (DE) [1] is a special case of these GAs, which has been proved to solve a broader range of optimization problems, thus, attracting even more interests in the last few years. Another important advantage of DE besides its robustness is its simplicity. In fact, the original DE was just about 20 lines of code, considerably smaller compared to other GAs.

The algorithm proposed in this paper is a modification to DE that helps enhancing its search speed without influencing the final result's accuracy. The main idea is

based on a geometrical concept in which 2 vectors with components sorted the same way always have a smaller difference (norm of the difference), further details are provided in section 3.

The rest of the article is organized as follows: Section 2 gives a brief description of DE; the proposed method is detailed in section 3. Simulation settings and experimental results are presented in section 4 and the conclusion is drawn in section 5.

2 DE Description

Differential Evolution (DE) is a population-based stochastic search algorithm that encodes its elements during each generation towards a global optimum. The population size is generally denoted by NP and each individual in the population is a vector $\mathbf{X}_{i,G} = (x_{i,G}^1, x_{i,G}^2, \dots, x_{i,G}^D)$ of dimension D . As in biological evolution, DE's algorithm is mainly divided into three parts: Mutation, crossover and selection. Several variants of DE exist and are mainly different in the way mutation and crossover is performed. However, the variant considered in this paper is the most used and is called DE/rand/1/bin [1]. Since global optima are usually unknown before applying the algorithm, therefore the population should be initialized in the search space in a uniformly random manner.

Mutation

For each target vector $X_{i,G}$ in the current generation, a mutant vector $V_{i,G}$ is obtained by randomly choosing a vector $X_{r1,G}$ and adding it to the scaled difference of two randomly chosen vector $X_{r2,G}$ and $X_{r3,G}$ such that i, r_1, r_2, r_3 are all mutually different. The resulting mutant vector can be expressed as follows:

$$V_{i,G} = X_{r1,G} + F \times (X_{r2,G} - X_{r3,G}) \tag{1}$$

F is the scaling factor and is generally chosen between 0 and 1.

Crossover

Given a target vector $X_{i,G}$ and the corresponding mutant vector $V_{i,G}$, a trial vector $U_{i,G}$ is obtained by combining the elements (genes) of the former two vectors as follows:

$$u_{i,G}^j = \begin{cases} v_{i,G}^j, & \text{if } \text{rand}_j(0,1) \leq C_r \text{ or } j = j_{\text{rand}} \\ x_{i,G}^j & \text{otherwise} \end{cases} \tag{2}$$

The constant $C_r \in [0,1]$ is the crossover rate, $\text{rand}_j \in [0,1]$ is the j^{th} evaluation of an uniform random number generator in the interval $[0,1]$, $j_{\text{rand}} \in \{1,2,\dots,D\}$ is a random parameter index chosen once for each i .

Selection

At this stage a natural selection has to be performed to choose the individual yielding lower objective (cost) between the target and trial vectors, i.e.:

$$\mathbf{X}_{i,G+1} = \begin{cases} \mathbf{U}_{i,G} & \text{if } f(\mathbf{U}_{i,G}) < f(\mathbf{X}_{i,G}) \\ \mathbf{X}_{i,G} & \text{Otherwise} \end{cases} \quad (3)$$

The same procedure is usually repeated for a given amount of generations or until a certain cost has been reached. Various stopping criteria might be implemented depending on the optimization problem at stake.

3 Proposed Method: Gene Sorting

During the past few years, various variants of DE have been proposed to increase its robustness and convergence speed [2-4]. These variants mainly aim at (self-) adapting the scaling factor, crossover rate, the population size and/or the mutation scheme to the problem at stake instead of manually tuning them. In fact, the trial-and-error method can be very time consuming since the parameter selection is problem-dependent.

However, the method proposed in this article takes a different prospective and does not attempt to adapt the algorithm parameters, but simply suggests to sort or order the genes in trial vectors. Therefore, this method can be combined with other DE variants to further increase their respective search speeds.

3.1 Theoretical Justification

A definition and two theorems have to be enounced before presenting the proposed method.

Definition 1. Given two vectors $\mathbf{X}=(x_1,x_2,\dots,x_n)$, $\mathbf{Y}=(y_1,y_2,\dots,y_n)$ and a permutation (interleaver) π such that $\mathbf{X}'=(x'_1,x'_2,\dots,x'_n)=\pi(\mathbf{X})$, with $x'_i \leq x'_{i+1} \forall i$. Then \mathbf{X} and \mathbf{Y} are said to be ordered in the same way iff $y'_i \leq y'_{i+1} \forall i$ with $\mathbf{Y}'=(y'_1,y'_2,\dots,y'_n)=\pi(\mathbf{Y})$

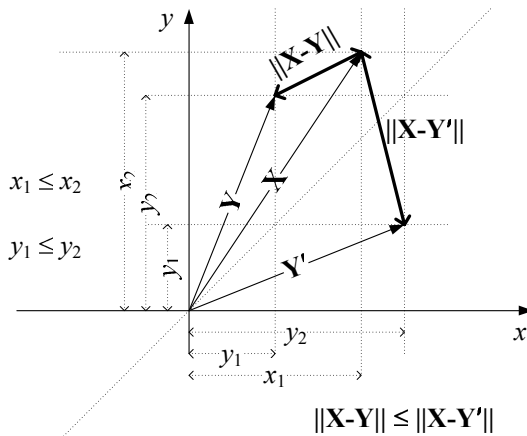


Fig. 1. Norm of the difference vector with and without sorted components in 2-dimension space

Theorem 1. *Given any vector $\mathbf{X}=(x_1,x_2,\dots,x_n)$ with sorted components, .i.e. $x_i \leq x_{i+1} \forall i$ and another vector $\mathbf{Y}=(y_1,y_2,\dots,y_n)$, then there is always less distance between \mathbf{X} and \mathbf{Y} if the components of \mathbf{Y} are also sorted in an ascending order as illustrated on Fig.1 (2-dimension case).*

Proof. The proof is divided into two parts and we supposed that vector $\mathbf{X}=(x_1,x_2,\dots,x_n)$ is sorted in an ascending order i.e. $x_i \leq x_{i+1} \forall i$.

a) Given a vector $\mathbf{Y}=(y_1,y_2,\dots,y_n)$ with at least two ordered elements $y_k \leq y_l$ with $k \leq l$ and a third vector \mathbf{Y}' with the same elements as \mathbf{Y} in the same order but with these two elements at interchanged position, i.e.:

$$\mathbf{Y}' = (y_1, y_2, \dots, y_{k-1}, y_l, y_{k+1}, \dots, y_{l-1}, y_k, y_{l+1}, \dots, y_n) . \tag{4}$$

Then

$$\begin{aligned} \|\mathbf{X}-\mathbf{Y}\|^2 - \|\mathbf{X}-\mathbf{Y}'\|^2 &= (y_k-x_k)^2+(y_l-x_l)^2 - (y_l-x_k)^2 - (y_k-x_l)^2 \\ &= -2(y_l-y_k)(x_l-x_k) \leq 0 . \end{aligned} \tag{5}$$

So

$$\|\mathbf{X}-\mathbf{Y}\| \leq \|\mathbf{X}-\mathbf{Y}'\| . \tag{6}$$

b) We consider a random vector $\mathbf{Y}=(y_1,y_2,\dots,y_n)$ then using a sorting algorithm as bubble sort [5] and because of a) the obtained vector at each iteration step of the algorithm is always closer to \mathbf{X} than the previous one, i.e.: $\|\mathbf{X}-\mathbf{Y}^{(i+1)}\| \leq \|\mathbf{X}-\mathbf{Y}^{(i)}\|$, the superscript denoting the iteration index. At the final step, we call $\mathbf{Y}_{\text{sorted}}$ the obtained sorted version of \mathbf{Y} and we conclude that

$$\|\mathbf{X}-\mathbf{Y}_{\text{sorted}}\| \leq \|\mathbf{X}-\mathbf{Y}\| . \tag{7}$$

■

Theorem 2. *The distance between two vectors (points) is always smaller if their coordinates are ordered in the same way.*

Proof. Given a vector $\mathbf{X}=(x_1,x_2,\dots,x_n)$ with its elements in a random order and a permutation (interleaver) π such that:

$$\mathbf{X}' = (x'_1,x'_2,\dots,x'_n) = \pi(\mathbf{X}), \text{ with } x'_i \leq x'_{i+1} \forall i$$

The permutation is bijective and therefore the inverse operation π^{-1} exists. For any random vector \mathbf{Y} and its image \mathbf{Y}' with respect to π .

$$\begin{aligned} \|\mathbf{X}-\mathbf{Y}\| &= \|\pi^{-1}(\mathbf{X}') - \pi^{-1}(\mathbf{Y}')\| \\ &= \|\pi^{-1}(\mathbf{X}' - \mathbf{Y}')\| \\ &= \|\mathbf{X}' - \mathbf{Y}'\| \end{aligned} . \tag{8}$$

It can be deduced that when the elements of \mathbf{Y} are given, then distance between \mathbf{X}' and \mathbf{Y}' is minimum when the elements of \mathbf{Y}' are ordered in an ascending order according to theorem 1, equivalently, the elements of \mathbf{Y} should be ordered in the same way as those of \mathbf{X} to yield a minimum difference.

3.2 Description of DE with Gene Sorting (DE_GS)

Following the previous observations, let the optimum solution be $\mathbf{X}^* = (x_1^*, x_2^*, \dots, x_n^*)$, therefore any trial vector $U_{i,G}$ is closer to \mathbf{X}^* if its components are ordered the same way as \mathbf{X}^* . In particular if the objective function is insensitive to the order of the vector's component, then for simplicity purpose, an ascending order might be assumed by default. Notice that the rest of the DE algorithm is kept unchanged and left as is and the only modifications are done on the trial vector just after the mutation procedure. Therefore, the modifications proposed in this article are minor but the convergence speed is at least doubled in most of the cases.

The justification for the appellation "gene sorting" is because a trial vector is equivalent to a chromosome and therefore its components are genes. Sorting them is equivalent to "sorting genes", hence this procedure can also be seen in a biological aspect as gene specialization or specification. Hereafter, DE with Gene sorting will then be called DE_GS.

However, DE_GS should not be applied when there is absolutely no information or statistics on the optimum solution. In this case, the default DE algorithm (or the corresponding variants) should be used instead.

The proposed algorithm can be further optimized in the case the cost function is insensitive to component ordering. In this specific case, trial vectors can be ordered only if they have successfully passed the natural selection step. This can help avoiding unnecessary ordering for unsuccessful trial vectors.

4 Experimental Results

The experiments have been conducted on a benchmark set of 18 functions selected from [3, 4] and all of dimension equal or greater than 10 defined as in Table 1. The choice of higher dimensional problems is because they are more challenging as far as speed is concerned. The stopping criteria are when the best individual in the population reaches a certain minimum called the Value-To-Reach (VTR) or when the maximum number of function evaluation, MAXnfe, is reached. During the experiments, VTR is set to $f(x^*) + 1E-8$ for all functions except for f_{14} , f_{16} and f_{18} where it is set to $f_{14}(x^*) - 1E-8$ (This is a maximization problem), $f_{16}(x^*) + 0.1$ and $f_{18}(x^*) + 6.7E-4$ respectively. In order to reduce the effect of stochastic hazard, each experiment in this section is run 50 times and only the average result is presented.

The benchmark functions are listed in Table 1, their respective dimension, search range and minima are also given:

Table 1. Benchmark functions: Sphere(f_1), Axis parallel hyperellipsoid(f_2), Schwefel’s problem 1.2(f_3), Rosenbrock(f_4), Rastrigin(f_5), Griewank(f_6), Sum of different power(f_7), Ackley(f_8), Levy(f_9), Zakharov(f_{10}), Schwefel’s problem 2.22(f_{11}), Step(f_{12}), Alpine(f_{13}), Exponential(f_{14}), Paviani(f_{15}), Salomon(f_{16}), De jong’s function 4(no noise) (f_{17}), Schwefel(f_{18}).

Test Function	D	S	x^*	f_{\min}
$f_1(x) = \sum_{i=1}^D x_i^2$	30	$[-5.12, 5.12]^D$	$(0,0,\dots,0)$	0
$f_2(x) = \sum_{i=1}^D ix_i^2$	30	$[-5.12, 5.12]^D$	$(0,0,\dots,0)$	0
$f_3(x) = \sum_{i=1}^D (\sum_{j=1}^i x_j)^2$	20	$[-65, 65]^D$	$(0,0,\dots,0)$	0
$f_4(x) = \sum_{i=1}^{D-1} [100(x_{i+1} - x_i)^2 + (1 - x_i)^2]$	30	$[-2, 2]^D$	$(1,1,\dots,1)$	0
$f_5(x) = 10D + \sum_{i=1}^D (x_i^2 - 10 \cos(2\pi x_i))$	10	$[-5.12, 5.12]^D$	$(0,0,\dots,0)$	0
$f_6(x) = \sum_{i=1}^D x_i^2 / 4000 - \prod_{i=1}^D \cos(x_i / \sqrt{i}) + 1$	30	$[-600, 600]^D$	$(0,0,\dots,0)$	0
$f_7(x) = \sum_{i=1}^D x_i ^{i+1}$	30	$[-1, 1]^D$	$(0,0,\dots,0)$	0
$f_8(x) = -20 \exp(-0.2 \sqrt{\sum_{i=1}^D x_i^2 / D}) - \exp(\sum_{i=1}^D \cos(2\pi x_i) / D) + 20 + e$	30	$[-32, 32]^D$	$(0,0,\dots,0)$	0
$f_9(x) = \sin^2(3\pi x_1) + \sum_{i=1}^{D-1} (x_i - 1)^2 \times (1 + \sin^2(3\pi x_{i+1})) + (x_D - 1)^2 (1 + \sin^2(2\pi x_n))$	30	$[-10, 10]^D$	$(1,1,\dots,1)$	0
$f_{10}(x) = \sum_{i=1}^D x_i^2 + (\sum_{i=1}^D 0.5ix_i)^2 + (\sum_{i=1}^D 0.5ix_i)^4$	30	$[-5, 10]^D$	$(0,0,\dots,0)$	0
$f_{11}(x) = \sum_{i=1}^D x_i + \prod_{i=1}^D x_i $	30	$[-10, 10]^D$	$(0,0,\dots,0)$	0
$f_{12}(x) = \sum_{i=1}^D \lfloor x_i + 0.5 \rfloor^2$	30	$[-100, 100]^D$	$0.5 \leq x_i \leq 0.5$	0
$f_{13}(x) = \sum_{i=1}^D x_i \sin(x_i) + 0.1x_i $	30	$[-10, 10]^D$	$(0,0,\dots,0)$	0
$f_{14}(x) = \exp(-0.5 \sum_{i=1}^D x_i^2)$	10	$[-1, 1]^D$	$(0,0,\dots,0)$	1
$f_{15}(x) = \sum_{i=1}^D [(\ln(x_i - 2))^2 + (\ln(10 - x_i))^2] - (\prod_{i=1}^D x_i)^{0.2}$	10	$[2, 10]^D$	$x_i=9.351$	-45.778
$f_{16}(x) = 1 - \cos(2\pi \ x\) + 0.1 \ x\ $	10	$[-100, 100]^D$	$(0,0,\dots,0)$	0
$f_{17}(x) = \sum_{i=1}^D ix_i^4$	30	$[-1.28, 1.28]^D$	$(0,0,\dots,0)$	0
$f_{18}(x) = 418.9829 \times D - \sum_{i=1}^D x_i \sin(x_i ^{1/2})$	30	$[-500, 500]^D$	420.969	0

4.1 Experiment Series 1: Comparison between DE and DE_GS with Settings from [4]

An attempt is made to empirically compared DE to DE_GS based on their average success performance (SP). For sake of comparability, most of the settings used here are the same as those in [4]. SP is a metric that takes into account the average number of function evaluations (nfe) and the average success rate (SR) defined as follows:

$$SR = \frac{\text{number of times VTR was reached}}{\text{total number of trials}}, \quad SP = \frac{\text{mean(nfe for successful runs)}}{SR} \tag{9}$$

The acceleration rate is also defined as:

$$AR = \frac{SP_{DE}}{SP_{DE_GS}} \quad (10)$$

AR is a metric that shows how fast DE_GS is, compared to DE. If $AR \geq 1$ then DE_GS is faster, otherwise it is slower.

Parameter settings [4]

- Population size, $NP = 100$
- Scaling factor $F = 0.5$
- Crossover rate $C_r = 0.9$
- Maximum number of function evaluations $MAX_{nfc} = 1E+6$

The comparison result is shown in table 2.

Table 2. Comparison between DE and DE_GS for 18 test functions

	D	DE	DE_GS	AR
f_1	30	85692	58128	1.474195
f_2	30	94838	45950	2.063939
f_3	20	168066	22598	7.437207
f_4	30	412508	239608	1.721595
f_5	10	329762	22718	14.51545
f_6	30	111414	-	-
f_7	30	28496	7581	3.758871
f_8	30	177960	104173	1.708312
f_9	30	95042	-	-
f_{10}	30	389580	50904	7.65323
f_{11}	30	184670	34408	5.367066
f_{12}	30	34600	23907	1.447275
f_{13}	30	372468	39458	9.439607
f_{14}	10	19590	9875	1.983797
f_{15}	10	15638	8171	1.913842
f_{16}	10	36930	8990	4.107898
f_{17}	30	50021	48662	1.027927
f_{18}	30	-	28400	-
Average		160055	48342	4.374681

According to Table 4, DE_GS is in average more than 4 times faster than DE and can get 7 to 9 times faster for some functions. Hence, gene sorting can clearly increase the convergence speed of DE. However, similarly to DE, DE_GS also seems to be influenced by the parameter settings (F and C_r). In fact, DE did not find the optimum for f_{18} , the same happened to DE_GS for f_{16} and f_{19} (it is shown by the *dashes* in the table).

Therefore, the comparison shown in Table 2 could not be considered complete if not tested for values of F and C_r with which each algorithm performs the best, respectively.

4.2 Experiment Series 2: Comparison of DE and DE_GS with Their Respective Best Control Parameters

The purpose of this experiment is to perform a “fair” comparison between DE and DE_GS. Therefore, only their best performances are compared in this subsection. The parameters for which DE and DE_GS respectively perform the best are found by keeping the dimension and the population size intact as in the previous experiment, but varying C_r between 0 and 1, F between 0 and 1.5 with a precision of 0.1. Since, each simulation is run for 50 trials, the best settings are found after running $50 \times 11 \times 16 = 8800$ trials for each function. Hence, this operation is extremely time-consuming; therefore, only 7 functions were chosen for the procedure (including those for which DE or DE_GS did not convergence in the previous experiment): $f_1, f_6, f_7, f_9, f_{10}, f_{13}$ and f_{18} . Similar results should be obtained for the other functions. The settings for which they perform the best and the corresponding acceleration rates are shown in Table 3.

Table 3. The best parameter settings and AR for $f_1, f_6, f_7, f_9, f_{10}, f_{13}$ and f_{18}

Test function	DE			DE_GS			AR
	C_r	F	$min(SP)$	C_r	F	$min(SP)$	
f_1	0.6	0.2	29390	1.0	0.6	10788	2.72
f_6	0.5	0.2	37651	1.0	0.7	17530	2.15
f_7	0.6	0.1	8149	1.0	0.5	3438	2.37
f_9	0.6	0.2	31352	1.0	0.6	13349	2.35
f_{10}	0.9	0.4	224638	0.9	0.6	36002	6.24
f_{13}	0.9	0.3	93530	1.0	0.6	27418	3.41
f_{18}	0.1	0.1	44248	1.0	0.8	17820	2.48

It is clear from these results that DE_GS is in average 3 times faster than classic DE (Note that NP is kept to 100) for their respective best control parameters.

In the meantime, from Table 3 and other results not shown here, it seems that 1.0 and 0.6 could be good initial choices for C_r and F respectively, for DE_GS, regardless of the function to be minimized. But further investigations need to be conducted in this regard. However, the integration of gene sorting should also be easy and straightforward in variants of DE with adaptive or self-adaptive schemes.

The last experiment of these series is the comparison between DE and DE_GS while the population size varies between $5 \times D$ and $10 \times D$ (Population range proposed in [1]) for their best settings (as shown in Table 3). The results are shown in Table 4 for $D=10, D=30$ and $D=60$ respectively.

Table 4 shows that DE_GS is in average always faster compared to DE and it gets faster as the problem dimension increases. The average acceleration rates are 2.0, 4.7 and 11.4 for 10-dimension, 30-dimension and 60-dimension ($VTR=1e-3$ for f_{18}) problems, respectively. Hence, DE’s speed could be considerably increased for high dimensional problems.

Table 4. Acceleration rates for $f_1, f_6, f_7, f_9, f_{10}, f_{13}$ and f_{18} with the best settings for DE and DE_GS. $D = 10, 30, 60$; $NP = 5xD, 6xD, 7xD, 8xD, 9xD, 10xD$

	f_1	f_6	f_7	f_9	f_{10}	f_{13}	f_{18}
<i>NP</i>	<i>D=10</i>						
5xD	1.76	2.82	4.29	1.50	5.06	0.70	1.29
6xD	0.43	2.99	3.43	1.29	2.11	2.03	1.13
7xD	1.69	2.12	3.06	1.46	1.92	2.35	1.08
8xD	1.67	2.07	2.10	1.43	1.94	2.67	1.13
9xD	1.71	1.98	1.63	1.43	1.97	2.85	1.15
10xD	1.69	2.32	1.67	1.47	2.05	3.31	1.13
Mean AR = 2.00							
<i>NP</i>	<i>D = 30</i>						
5xD	2.89	2.18	2.55	2.47	6.83	4.96	2.38
6xD	2.92	2.30	2.41	2.49	7.48	7.07	2.42
7xD	2.89	2.35	2.43	2.55	7.87	9.04	2.43
8xD	2.93	2.26	2.40	2.53	8.17	14.63	2.41
9xD	2.97	2.30	2.50	2.54	8.66	17.49	2.44
10xD	2.99	2.34	2.44	2.58	9.07	21.90	2.42
Mean	Mean AR = 4.73						
<i>NP</i>	<i>D = 60</i>						
5xD	4.23	3.23	3.42	3.69	22.17	7.65	3.63
6xD	4.30	3.27	3.54	3.78	24.88	13.22	3.66
7xD	4.27	3.28	3.52	3.76	26.85	17.51	3.66
8xD	4.36	3.26	3.41	3.78	28.83	36.35	3.68
9xD	4.41	3.30	3.71	3.79	30.52	46.80	3.66
10xD	4.46	3.29	3.64	3.82	31.92	81.76	3.66
Mean	Mean AR = 11.43						

4.3 Experiment Series 3: Application of Gene Sorting on a Variant of DE Called Iterated Function System Based Adaptive Differential Evolution (IFDE)

A detailed presentation of IFDE can be found in [6] and IFDE with gene sorting is called IFDE_GS hereafter. The comparison results are shown in Fig.2, Fig.3, Fig.4 and Fig.5 (not all the functions could be shown here due to space limitation). Mean Value represents the average minimum value (over 50 trials) obtained at each stage of the generation.

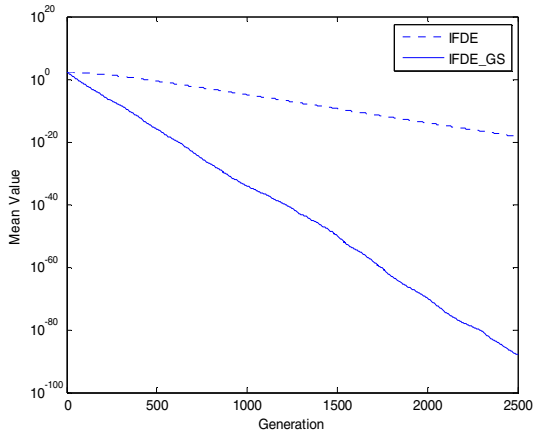


Fig. 2. IFDE and IFDE_GS comparison on the sphere function

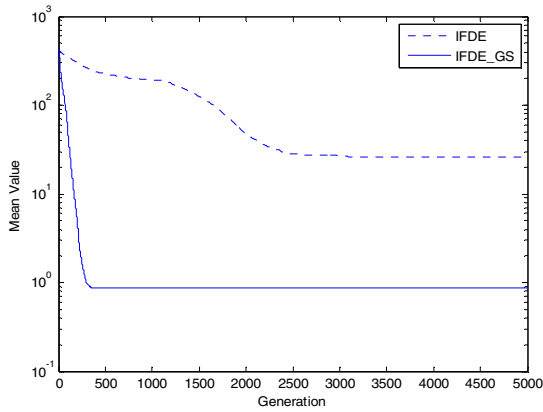


Fig. 3. IFDE and IFDE_GS comparison on the Rastrigin's function

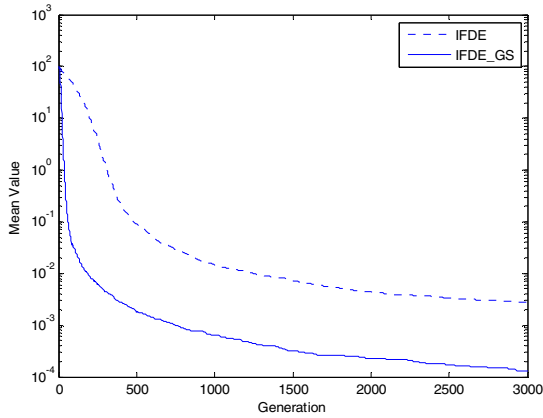


Fig. 4. IFDE and IFDE_GS comparison on the Schwefel's problem 1.2

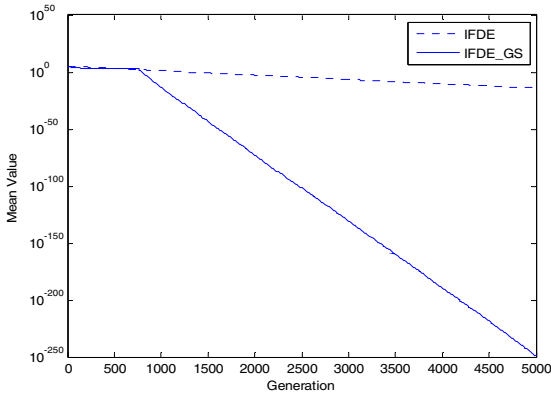


Fig. 5. IFDE and IFDE_GS comparison on the De Jong's function with noise

From these figures, it can be seen that with gene sorting, IFDE is 5 times faster for the sphere's function and the De Jong's function with noise, 6 times faster for Schwefel's problem 1.2 and 10 times faster for Rastrigin's problem.

In view of these results, we can conclude that the combination of gene sorting to other variants of DE could bring promising outcomes and it is worth further investigations.

5 Conclusion

Differential evolution is a simple, yet efficient genetic algorithm for global optimization problems that has gained an increased attention over the past few years in a broad range of research fields. In this paper, we show from a geometric point of view that ordering the components of trial vectors (gene sorting) in the same order as the optimum solution can very much increase the convergence speed of the algorithm. Even though optimum solutions are unknown prior to search, the component's order of the optimum point for various objective functions can be deduced prior to search. A special case is when the objective function is insensitive to component's ordering; in this particular case, these components can be sorted in an ascending order by default. We also suggest the application of gene sorting in other variants of differential evolution including those with adaptive and/or self-adaptive schemes to further increase their respective convergence speeds.

Acknowledgments. Corresponding Author: Desheng Wang, dswang@hust.edu.cn This work is partly supported by the National Natural Science Foundation of China under Grant No.60802009 and No.60496315, the International Science and Technology Cooperation Programme of China under Grant No.2008DFA11630 and No.2008AA01Z204, Hubei Science Foundation under Grant No.2007ABA008, and Postdoctoral Foundation under Grant No.20070410279. The authors would like to thank Shahryar, R. and Li, Y.L. for the invaluable help they brought during the writing of this article.

References

1. Storn, R., Price, K.: Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *Journal of Global Optimization* 11 (1997)
2. Junhong, L., Jouni, L.: A Fuzzy Adaptive Differential Evolution Algorithm. In: *TENCON 2002: Proceedings of the IEEE Region 10 Conference on Computers, Communications, Control and Power Engineering*, vol. 1, pp. 606–611 (2002)
3. Qin, A.K., Huang, V.L., Suganthan, P.N.: Differential Evolution Algorithm With Strategy Adaptation for Global Numerical Optimization. *IEEE Transactions on Evolutionary Computation* (2008)
4. Rahnamayan, S., Tizhoosh, H.R., Salama, M.M.A.: Opposition-Based Differential Evolution. *IEEE Transactions on Evolutionary Computation* 12, 64–79 (2008)
5. Astrachan, O.: Bubble Sort: An Archaeological Algorithm Analysis. In: *Technical Symposium on Computer Science Education*, Nevada, USA (2003)
6. Ya, L.L., Fei, D., Wang, Y.-X.: Iterated Function System Based Adaptive Differential Evolution Algorithm. In: *CEC 2008: IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)*, pp. 1290–1294 (2008)

Enhancement of Chest Radiograph Based on Wavelet Transform

Zhenghao Shi^{1,2}, Lifeng He³, Tsuyoshi Nakamura¹, and Hidenori Itoh¹

¹ School of Computer Science and Engineering, Nagoya Institute of Technology, Japan

² School of Computer Science and Engineering, Xi'an University of Technology, China

³ School of Information Science and Technology, Aichi Prefectural University, Japan
{tnaka, itoh}@juno.ics.nitech.ac.jp, ylshi@xaut.edu.cn,
helifeng@ist.aichi-pu.ac.jp

Abstract. The effect of anatomical noise is one of the major challenges for the early detection of pulmonary nodules in chest radiograph. A method aimed at eliminating these anatomical noises while enhancing contrast of anatomical feature is presented. The method is based on local modification of gradient magnitude values provided by the redundant dyadic wavelet transform. It includes two key steps. The first one is to threshold wavelet coefficients, which is accomplished by using a threshold strategy. The purpose of this operation is to reduce the effect of background and anatomical noise on the region of interesting in the chest radiograph. The second one is to do a normalization operation for all retained wavelet coefficients at a same scale. The purpose of this operation is to ensure that the enhanced image is not sensitive to the variance of radiograph acquirement environment. Experimental results (performed under different conditions.) indicate the efficiency and the effectiveness of the proposed method in radiography enhancement.

Keywords: Radiography enhancement, Anatomical noise, Wavelet transformation.

1 Introduction

Lung cancer is one of the most frequently occurring cancers of all cancer diagnoses, and detecting the disease in its early stages is the most promising strategy to enhance the survival chances of patients [1]. Chest radiography is currently the most common tool for the initial detection and diagnosis of lung cancer [2]. However, the early detection of pulmonary nodules in chest radiography images can be one of the most challenging in clinical practice [3]. The main reason is that a chest radiograph is a projection image which contains noise-like components because of the overlapping of fine anatomical structures. These components can not be distinguished as anatomical structures and thus referred as anatomical noises [4].

Our purpose in this study is to develop a technique that can be used to remove the anatomical noise from a chest radiograph. We employed wavelet transformation (WT) in our work. The advantage of this multi-scale technique is that the structures of different size appear at different scales and can be processed independently. Since we can express the original image as a combination of its approximations and different

levels of details, we can build an enhancement algorithm in WT domain. Special attention was paid when designing the enhancement function and the processing strategy for the finest scales. Experimental results (performed under different conditions.) indicate that the proposed method is capable of considerably improve the subjective image quality without providing any noticeable artifact.

The remainder of this paper is organized as follows: Section 2 describes the database used for this study. A simple description of wavelet transformation is given in Section 3 firstly, and then the method used in this study is depicted in detail. Section 4 shows the experimental results. Conclusions of this paper are shown in Section 5.

2 Database

The database used for this study included 91 chest radiographs containing 75 solitary pulmonary nodules and 410 non-nodules, selected from the Japanese Standard Digital Image (JSRT) Database developed by the Japanese Society of Radiological Technology, which is available publicly. The JSRT Database includes 154 abnormal chest radiographs, each with a solitary pulmonary nodule, and 93 non-nodule chest radiographs. These original screen-film images were digitized with a 0.175 mm pixel size, matrix size of 2048×2048 pixels, and 12 bits of gray scale. For increasing computational efficiency, the size of the chest radiographs was reduced to 500 by 500 pixels with a 10 bit gray scale level through the use of averaging from that of 2048 by 2048 with a 12 bit gray scale level.

3 Method

3.1 Wavelet Transformation [5]

In medical image processing, we usually deal with discrete data. We will therefore focus our discussion on discrete wavelet transform rather than continuous ones here.

Let

$$C(j, k) = (f(t) \bullet \psi_{j,k}(t)) \quad (1)$$

be the discrete wavelet transform (DWT) coefficients of signal $f(t)$ and let $\psi_{j,k}(t)$ be an orthogonal wavelet function. Given the details D (high-frequency information) and approximations A (low-frequency information) of the signal $f(t)$ at level j , reconstruction of the signal $f(t)$ from its WT coefficients at different scales is defined as

$$\begin{aligned} D_j(t) &= \sum_{k \in Z} C(j, k) \psi_{j,k}(t) \\ f(t) &= \sum_{k \in Z} D_j \\ A_j &= \sum_{j > J} D_j \\ A_{j-1} &= A_j + D_j \end{aligned} \quad (2)$$

Then the wavelet decomposition can be done by using iterated two-channel filter banks (as shown in figure 1).

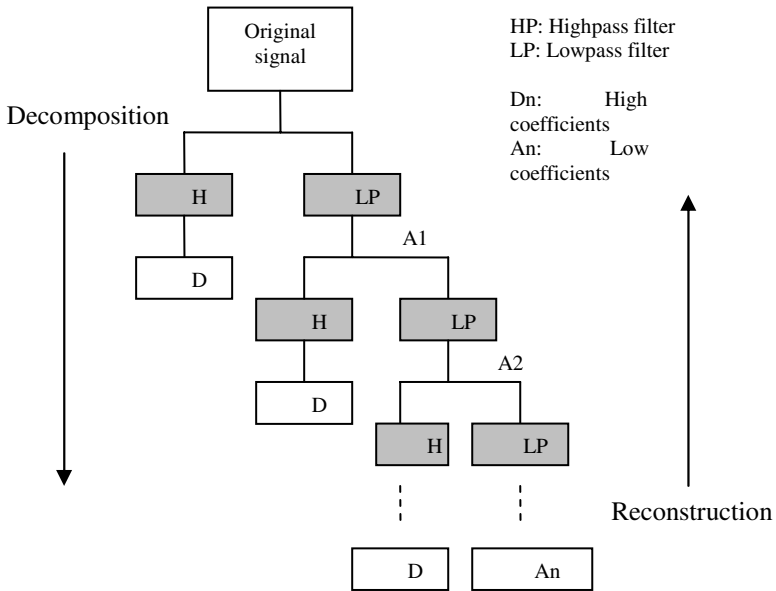


Fig. 1. Illustration of wavelet transforms. Wavelet transformation can be done by using iterated two-channel filter banks: high-pass filters and low-pass filters. The high-pass filters are detail coefficients, and the low-pass filter are approximation coefficients.

The down-sampled outputs of the high-pass filter are detail coefficients, and the down-sampled outputs of the low-pass filter are approximation coefficients. The detail and approximation coefficients provide an exact representation of the signal. Decomposing the approximation coefficients perform a further level of the detail and approximation coefficients.

The reconstruction process is done by inverse iterative two-channel filter bank, consisting of up-sampling from each channel, performing a synthesis low-pass and high-pass filtering, and summing up the results from both channels.

3.2 Radiography Enhancement by Wavelet Transformation

The mainly cause which leads to variance of pixel in chest radiograph is that the differences of attenuation coefficients between signals and background substance are subtle in some cases [6]-[8]. In these cases, small wavelet coefficients often appear in the slowly varying places, whereas large wavelet coefficients often appear in rapidly varying regions. Most structures of our interest often lie in slowly varying regions and have small wavelet coefficients since their attenuation coefficients differ slightly from those of the background [8].

Under this consideration, to enhance radiograph, we suggest that a non-linear mapping should be performed on wavelet coefficients. Our ideas are illustrated in figure 2, which consists of the following three operations:

Firstly, the original image is decomposed by WT.

Then contrast enhancement is done by modifying the coefficients obtained from the decomposed image. Small coefficient values represent subtle details and are amplified to improve the information of the corresponding details. The stronger image density variations make a major contribution to the overall dynamic range, and have large coefficient values. These values can be reduced without much information loss.

Finally, the inverse WT is done to obtain the enhanced image. To compute each lower level of approximation coefficients using the modified detail and approximation coefficients of the previous level of reconstruction until the final enhanced image is computed.

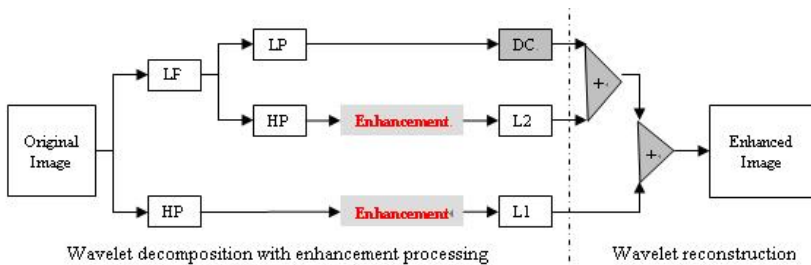


Fig. 2. Image enhancement based on wavelet transformation. Contrast enhancement is done by modifying the coefficients obtained from the decomposed image. Small coefficients are amplified to improve the information of the corresponding details. Large coefficient values are reduced without much information loss.

To modify the WT coefficients for obtaining a right enhanced image, the following basic constraints [9] should be considered and followed:

- (1) An area of low contrast should be enhanced more than an area of high contrast. This is equivalent to say that smaller values of wavelet coefficients should be assigned larger gains.
- (2) The radiograph enhancement must not cause misleading information, making a structure looking more or less significant than it is must be avoided.
- (3) Monotonically increasing: Monotonicity ensures the preservation of the relative strength of signal variations, and avoids changing location of local extremers, or creating new extremers.

Based on what mentioned above, the method for modification of WT coefficients and computation of the enhanced images from the modified transform coefficients is described in the following.

- (1). To modify the detail coefficients

A simple piecewise power transform [8] can be used here:

$$y(i, j) = \text{sgn}(x(i, j)) |x(i, j)|^\gamma, 0 < \gamma < 1 \quad (3)$$

where $x(i, j)$ is detail coefficients after first wavelet decompose), γ is a tunable parameter for controlling the enhancement level. Small γ provides higher enhancement and the enhanced image converges to the original image when γ approaches one.

Such enhancement is simple to implement, and was used successfully for contrast enhancement on mammograms [9]. However Further researches indicate that such simple enhancement operator is limited to contrast enhancement of data with very low noise level for the reason that this transform may also amplify noise components which related to wavelet coefficients with small magnitude [10].

Aim at this problem, in our study, a method proposed by Xu [10] is employed, where the correlation degree of wavelet coefficient between different neighborhood scales is computed first, where W_n is the wavelet coefficient of scale n, and W_{n+1} is the wavelet coefficient of scale n+1.

$$C_n = (W_n \times W_{n+1}) \times \sqrt{\frac{\sum (W_n \times W_n)}{\sum (W_n \times W_{n+1})^2}} \tag{4}$$

then who (noise or detail) produces the wavelet coefficients is determined according with the correlation degree values. If the absolute value of the wavelet correlation degree of a pixel is larger than that of the wavelet coefficient value in this pixel, then the wavelet coefficient of the pixel is produced by detail image signal. Otherwise the wavelet coefficient is produced by noise.

Next different processing is done to the wavelet coefficients produce by noise or details respectively. To avoid loss of weakly contrasting detail information, we refrain from integrating genuine delousing into our approach just by wavelet coefficient shrinkage. Instead, we seek to prevent unacceptable noise amplification, also to avoid creating any new discontinuities, during image enhancement. As a result, the following sigmoid function [11] is adopted to modify the detail wavelet coefficients in our study:

$$E(x(i, j)) = a[\text{sigm}(c(x(i, j) - b)) - \text{sigm}(-c(x(i, j) + b))] \tag{5}$$

where

$$a = \frac{1}{\text{sigm}(c(1 - b)) - \text{sigm}(-c(1 + b))} \tag{6}$$

$$\text{sigm}(t) = \frac{1}{1 + e^{-t}} \tag{7}$$

the parameters b ($0 < b < 1$) and c respectively control the threshold and rate of enhancement. It can be easily seen that $E(x)$ in Equation (7) is continuous and monotonically increasing within the interval $[1, 1]$. Furthermore, any order of derivatives of $E(x)$ is existent and continuous. This property avoids creating any new discontinuities after enhancement.

(2). To attenuate the approximation coefficients

The human visual system is sensitive to the different spatial frequencies in an image. In particular, the plots for human visual frequency indicate that some frequencies are more visible than the others, and some are not important at all. Removing certain frequencies can help emphasize the others (keeping the total image energy the same), and improving the quality of the image [12].

For image wavelet transform, because the stronger image density variations make a major contribution to the overall dynamic range, so its have large coefficient values. Based on the principle of human visual system mentioned above, these values can be reduced without much information loss. For this purpose, here the following equation is adopted.

$$A'(i, j) = \alpha A(i, j), 0 < \alpha < 1 \tag{8}$$

(3). To normalize wavelet coefficients

In fact, with the variance of wavelet coefficients, the absolute pixel values of a radiograph have also been changed. In order keep the image information energy scale invariance, when all operations mentioned above are finished, we will do a normalization operation for all different wavelet coefficients at a same scale, as show in Equation (9). This operation ensures the enhanced image has a same energy scale with its corresponding original image. The normalized image may be not sensitive to the variance of radiograph acquirement environment and may be better for further used in a computer aided diagnosis system.

$$E(I(x, y)) = \frac{E(I(x, y))}{\sum_{x=1}^M \sum_{y=1}^N E(I(x, y)) / (M * N)} \tag{9}$$

4 Experiments

In this section an experimental validation of the proposed method is provided. In our experiment, the performance of the proposed method is evaluated via its effect on ROI images and its corresponding output ROI images by our current CAD system (4.1) and by its influence on the performance of our CAD in the false positive reduction of nodule detection in chest radiograph (4.2).

4.1 Effect of the Proposed Method on ROI Images and Its Output ROI Images

The anatomical noise removing method based on wavelet transform was applied to our database of 91 original chest radiographs which contains 75 nodules and 410 non-nodules. Figure 3 illustrates the processed ROI images without and with wavelet transform and their corresponding output ROI images by our CAD system, respectively. From Figure 3, it can be seen that the boundary of ROI images without wavelet transformation are a little blur in visual because of the anatomical noise, while the ones with wavelet transformation are very clearly. The output ROI images by our CAD system with wavelet transformation are lighter than that of ones without wavelet

transformation. Compared with Fig.3 (a), Fig.3 (b) reveals fine details. This indexes that the wavelet anatomical noise removing method is helpful for improve the performance of our current computer aided nodules detection scheme.

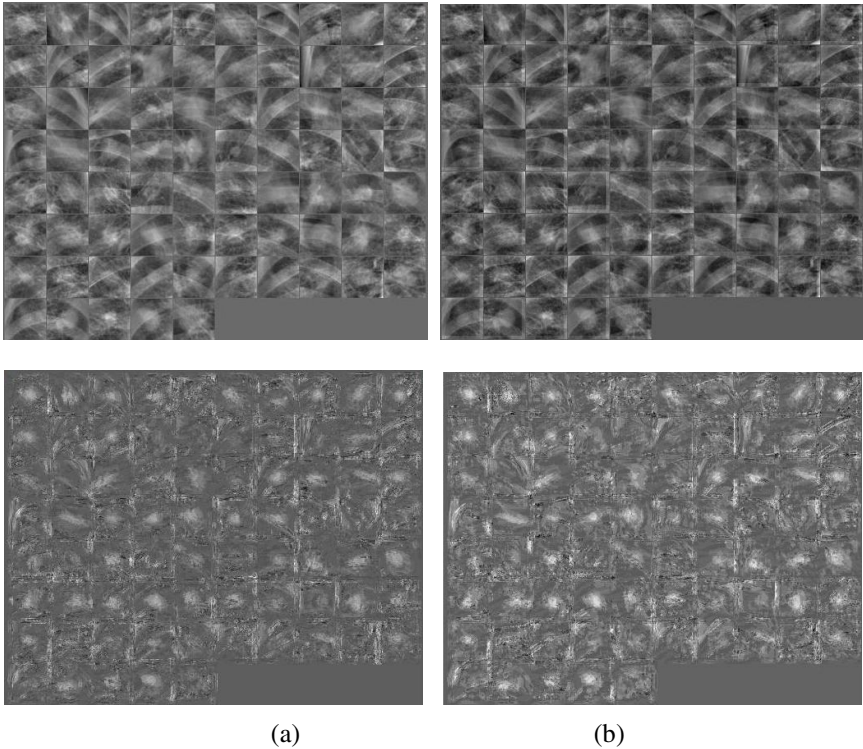


Fig. 3. ROI images without and with wavelet enhancement. (a) ROI images without wavelet enhancement and its corresponding output ROI images by our CAD. (b) ROI images with wavelet enhancement and its corresponding output ROI images by our CAD.

4.2 Influence of Proposed Method on Our Current CAD in False Positive Reduction of Nodule Detection in Chest Radiograph

In order to investigate the influence of the proposed method on the performance of our CAD system in false positive reduction of nodule detection in chest radiograph, we applied our CAD system to the 75 true positives (nodules) and 410 false positives (non-nodules) under two conditions: (1) without anatomical noise remove with wavelet transformation, (2) removing anatomical noises by wavelet transformation. The performance of our CAD system was evaluated by FROC (Free Response Receiver Operating Characteristic) curve. The FROC curve expresses the sensitivity as a function of the number of false positives per image at a specific operating point on the curve. Figure 4 shows the FROC curves for our CAD under the two conditions. With the anatomical noise elimination, the number of false positives of our current CAD

was reduced from 4.8 to 3.2 false positive per image, at an overall sensitivity of 86%. The results indexes that the proposed method is helpful for the improvement of performance of our CAD system in lung nodule detection in chest radiograph.

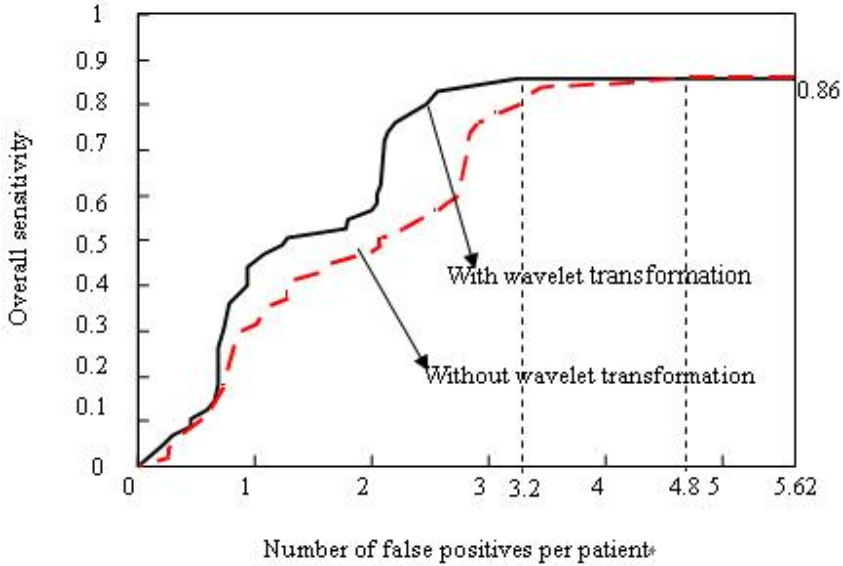


Fig. 4. FROC curve of (thick solid curve) of our CAD for 75 true positives (nodules) and 410 false positive (non-nodules) with wavelet transformation enhancement preprocessing, and FROC curve (dotted curve) of our CAD for the same database without wavelet transformation enhancement preprocessing

5 Conclusions

A novel method used for removing anatomical noise of chest radiograph is proposed, which is based on local modification of gradient magnitude values provided by the redundant dyadic wavelet transform. The advantage of this technique is that structures of different size appear separately in different scales, at each scale, weakly contrasting structures can be enhanced by suitable nonlinear processing independently. Experimental results demonstrate the efficiency and the effectiveness of the proposed method in the lung nodule detection in chest radiograph based on our current CAD.

Acknowledgments. The authors are grateful to the anonymous referees for their constructive and helpful comments.

References

1. Jemal, A., Siegel, R., Ward, E., Hao, Y.P., Xu, J.Q.: Taylor Murray and Michael J. Thun: Cancer statistics. CA: A Cancer Journal for Clinicians 58, 71–96 (2008)
2. Doi, K.: Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Computerized medical imaging and graphics 31, 198–211 (2007)

3. Ginneken, B.V., Bart, M., Romeny, H., Viergever, M.A.: Computer aided diagnosis in chest radiograph: a survey. *IEEE transaction on medical imaging* 12, 1228–1240 (2001)
4. Chen, S.Y., Hou, H.H., Zeng, Y.J., Xu, X.M.: Study of Automatic Enhancement for Chest Radiograph. *Journal of Digital Imaging* 4, 371–375 (2006)
5. Mallat, S.: A theory for multi-resolution signal decomposition: The wavelet representation. *IEEE transactions on pattern analysis and machine Intelligence* 7, 674–693 (1989)
6. Qi, Z.H., Zhang, L., Xing, Y.X., Li, S.L.: X-ray image enhancement based on the dyadic wavelet transform. *Journal of X-Ray Science and Technology* 14, 83–93 (2006)
7. Dippel, S., Stahl, M., Wiemker, R., Blaffert, T.: Multiscale contrast enhancement for radiographies: Laplacian pyramid versus fast wavelet transform. *IEEE Trans. Med. Imaging* 4, 343–353 (2002)
8. Hakan, O., Karen, E., Jarkko, N., Juha, L.: An Approach to Adaptive Enhancement of Diagnostic X-Ray Images. *EURASIP Journal on Applied Signal Processing* 5, 430–436 (2003)
9. Laine, A., Fan, J., Yang, W.: Wavelets for Contrast Enhancement of Digital Mammography. *IEEE Engineering in Medicine and Biology* 5, 536–550 (1995)
10. Xu, Y., Weaver, J.B., Healy, D.M., et al.: Wavelet transform domain filters: A spatially selective noise filtration technique. *IEEE Transactions on Image Processing* 6, 747–758 (1994)
11. Koren, I., Laine, A.: A discrete dyadic wavelet transform for multidimensional feature analysis in Time Frequency and Wavelets in Biomedical Signal Processing. In: Akay, M. (ed.) *IEEE Press series in biomedical engineering*, pp. 425–448. IEEE Press, Piscataway (1998)
12. Ji, T.L., Sundareshan, M.K., Roehrig, H.: Adaptive image contrast enhancement based on human visual properties. *IEEE Trans. Med. Imaging* 4, 573–586 (1994)

Application of DNA Computing by Self-assembly on 0-1 Knapsack Problem

Guangzhao Cui¹, Cuiling Li¹, Xuncaizhang^{1,2}, Yanfeng Wang¹, Xinbo Qi³,
Xiaoguang Li¹, and Haobin Li¹

¹ Henan Key Lab of Information-based Electrical Appliances, Zhengzhou 450002, China

² Department of Control Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China

³ Henan Mechanical and Electrical Engineering College, Xinxiang 153000, China
lilingcui@163.com

Abstract. Computation by self-assembly of DNA is an efficient method of executing parallel DNA computing where information is encoded in DNA tiles and a large number of tiles can be self-assembled via sticky-end associations. It presents clear evidence of the ability of molecular computing to perform complicated mathematical operations. We investigate how basic ideas on tiling can be applied to solving knapsack problem. It suggests that these procedures can be realized on the molecular scale through the medium of self-assembled DNA tiles. The potential of DNA computing by self-assembly for the knapsack problem is promising given the operational time complexity of $\mathcal{O}(n)$.

Keywords: DNA computing, Self-assembly, Knapsack problem, DNA tile.

1 Introduction

Since the seminal work of Adleman on the HPP [1], the field of DNA based computing has experienced a flowering growth and leaves us with a rich legacy. Given its vast parallelism and high-density storage, DNA computing approaches have been employed to solve many combinatorial optimization problems. However, most of these proposals implement the computation by performing a series of biochemical reactions on a set of DNA molecules, which require human intervention at each step. Thus, one difficulty with such methods for DNA computing is the number of laboratory procedures.

Winfrey et al. first proposed the idea of computation by self-assembled tiles. Computation and self-assembly are connected by the theory of tiling, of which Wang tiles [2] are a prime example. Because that DNA tiles can be more easily “programmed” to incorporate the constraints of a given problem, it is possible to exercise some degree of controlling over the appearance of considerable waste of material that belongs to the traditional approach, and therefore parallel computation can be enhanced by self-assembling process, where information is encoded in DNA tiles. The relation of DNA computation to self-assembling structures was developed in the mid-1990s, largely through the theoretical and experimental work of Winfrey et al [3], [4], Seeman [5]

Reif [6] and Rozenberg et al. [7] Computational systems based on self-assembly have been demonstrated in both 1-D arrangements called “string tiles” [3], [8] and 2-D lattices of DNA [4]. Other stable forms of nucleic acids include Z-DNA, non-migrating Holliday junctions, and duplexes with triple crossovers or “pseudoknots” [5], [9], [10]. For 2-D self-assembly, Winfree has proposed the Tile Assembly Model (TAM)[4] and demonstrated that it is Turing- universal by showing that a tile system can simulate Wang tiles [2], which Robinson has shown to be universal [11].

The first experimental demonstrations of computation using DNA tile assembly was in Ref. [12]. It demonstrated a two-layer, linear assembly of triple-crossover (TX) tiles that executed a bit-wise cumulative XOR computation. Barish et al. demonstrated a DNA implementation of a tile system that copies an input and counts in binary [12]. Cook et al. used the TAM to implement arbitrary circuits [13]. Similarly, Rothmund et al. demonstrated a DNA implementation of a tile system that computes the XOR function, resulting in a Sierpinski triangle [14]. Brun proposed and studied some systems that compute the sums and products of two numbers using the TAM [15]. He then combined these systems to create systems with more complex behavior, and designed two systems that factors numbers and decides subset sum problems [16], [17]. We proposed two systems that compute the differences and quotients of two numbers [18].

2 DNA Self-assembly

DNA nanostructures provide a programmable methodology for bottom-up nanoscale construction of patterned structures, utilizing macromolecular building blocks called DNA tiles based on branched DNA. These tiles have sticky ends that match the sticky ends of other DNA tiles, facilitating further assembly into larger structures known as DNA tiling lattices.

2.1 DNA Tile

Branched DNA molecules provide a direct physical motivation for the TAM. There is also a logical equivalence between DNA sticky ends and Wang tile edges. These DNA tiles have unpaired ends of DNA strands sticking out and through these sticky ends, they can attach themselves with other tiles having the Watson-Crick complementary sticky end. Thus, these tiles can stick with one-another to assemble into complex superstructures and through this process they can compute in a way similar to Wang tiles.

The most relevant constructions for DNA computing by self-assembly of DNA tilings include the double-crossover (DX) [19] and triple- crossover (TX) [20] complexes. These tiles can be constructed using a variety of possible nucleotide sequences. We can use different sequences to denote different symbols or values. For example, we can have one sequence denoting the value 0 and another sequence denoting the value 1. We can also use the sticky ends of a tile to encode certain values or symbols. The tile that attaches to this tile would have the complementary sticky end encoding the same value or symbol. In this way, we can pass information from one tile to its adjoining tile.

2.2 Molecular Self-assembly Processes

Molecular Self-assembly is the ubiquitous process by which simple objects autonomously assemble into intricate complexes. It has been suggested that intricate self-assembly processes will ultimately be used in circuit fabrication, nanorobotics, DNA computation, and amorphous computing [21].

There are three basic steps that define a process of molecular self-assembly [22].

1. Molecular recognition: elementary molecules selectively bind to others;
2. Growth: elementary molecules or intermediate assemblies are the building blocks that bind to each other following a sequential or hierarchical assembly. Cooperativity and non-linear behavior often characterize this process;
3. Termination: a built-in halting feature is required to specify the completion of the assembly. Without it, assemblies can potentially grow infinitely; in practice, their growth is interrupted by physical and/or environmental constraints.

It has the following three characteristics [22]:

1. Molecular self-assembly is a time-dependent process and because of this, temporal information and kinetic control may play a role in the process before thermodynamic stability is reached.
2. Molecular self-assembly is also a highly parallel process, where many copies of different molecules bind simultaneously to form intermediate complexes.
3. Another characteristic of a molecular self-assembly is that the hierarchical build-up of complex assemblies allows one to intervene at each step, either to suppress the following one, or to orient the system towards a different pathway.

2.3 Programming Self-assembly of DNA Tiling

A tiling is an arrangement of a few tiles that fit together perfectly in the infinite plane. Programming DNA self-assembly of tilings amounts to the design of the pads of DNA tiles, ensuring that only the adjacent pads of neighboring tiles are complementary so that tiles assemble together as intended. The use of pads with complementary base sequences allows the neighbor relations of tiles in the final assembly to be intimately controlled; thus the only large-scale superstructures formed during assembly are those that encode valid mappings of input to output. The progress of self-assembly of DNA Tilings can be carried out by the following four steps: 1. mixing the input oligonucleotides to form the DNA tiles, 2. allowing the tiles to self-assemble into superstructures, 3. ligating strands that have been co-localized, 4. then performing a single separation to identify the correct output.

3 0-1 Knapsack Problem

0-1 knapsack problem is a special kind of 0-1 integer programming problem. This section will introduce 0-1 knapsack problem which displayed as formula (1). 0-1 knapsack problem is a typical NP-hard in the operation research field. Problems in cargo loading, cutting stock, budget control and financial management may be formulated as knapsack problems. So the research of an efficient algorithm will be significant both in theory and practice. Generally speaking, all the algorithms can be classed into two kinds. One can be called the exact algorithms, such as the dynamic programming and branch-and-bound method. The others are approximate ones, such as

the ant algorithm. Although most approaches for solving 0-1 knapsack problem were developed in recent years, no method can solve large instances owing to computational inefficiency.

$$\begin{aligned} & \text{Max} \sum a_i v_i \\ & \sum_{i=1}^n a_i w_i \leq b \quad a_i \in \{0, 1\} \end{aligned} \quad (1)$$

3.1 DNA Self-assembly for 0-1 Knapsack Problem

The aim of 0-1 knapsack problem is to search for a group of truth assignments of all the variables a_i ($i=1, 2, \dots, n$) which must satisfy the constraint inequalities of 0-1 knapsack problem, and then to evaluate the result of object function. To achieve this goal, Our idea is to exploit the massive parallelism possible in DNA self-assembly in order to solve the 0-1 knapsack problem in polynomial time. It is described at the algorithmic level.

We abstract each DNA tile as a square with labels at the edges. Each label indicates a particular kind of a sticky end. Two sticky ends that can match and ligate correspond to identical labels. Each tile can have any from 1 to 4 labels. Non-labeled edges indicate non-sticky ends. It is taken for granted here that a tile would not occupy a slot unless both labels match.

3.2 The Model of DNA Self-assembly for 0-1 Knapsack Problem

It is also possible that a set of tiles do not deterministically produce a unique tiling. In this case, there are many possible valid tilings, any or all of which may be produced. When tiles are implemented by real molecules, one would expect a set of nondeterministic tiles to generate a combinatorial library. This is exactly what is necessary to perform a massively parallel computation. As sketched in Fig.4 and Fig 7(b), the orange tiles, attaches nondeterministically, determining which tiles attach to its east in addition system (west in subtraction system), and then a deterministic set of rule tiles could evaluate each input assembly to determine whether it represents the desired answer.

Here, use the massive parallelism of DNA to check whether constraint inequalities is satisfied with all the possible inputs. Reading of the operation result is done by the reporter strand method [10]. We designed the following algorithm to solve 0-1 knapsack problem corresponding to formula (1):

Step 1: Generate all possible variables in the given problem;

Step 2: According to the rules for dealing with constraint in section 3.3 and 3.4, using the massive parallelism of DNA to check whether constraint inequality is satisfied with all the possible inputs.

Step 3: Reject all infeasible solutions according to constraint inequality (reserved feasible solution) and obtain feasible solutions of the problem;

Step 4: By comparing to value of object function corresponding to every feasible solution, we can obtain optimum solution.

3.3 L-Configuration Subtraction

First, we analyze the form of 0-1 knapsack problem. Given a vector $\bar{w} = (w_1, w_2, \dots, w_n)$, we need to calculate $\sum_{i=1}^n a_i w_i \leq b$ and it equal to $b - \sum_{i=1}^n a_i w_i \geq 0$. Here we consider using b subtract $\sum_{i=1}^n a_i w_i$ to judge the restriction condition. It similar to the method which has been described by Brun in Ref. [17]. Here we briefly introduce it, which contains 4 subsystems: 1. subtract computations. Its a system that subtracts positive integers and contains 16 tiles, it will subtract one bit per row of computation. Fig. 1 shows the 16 tiles of T_- . The value in the middle of each tile t represents that tile's $v(t)$ value. Intuitively, the system will subtract the i th bit on the i th row. The tiles to the right of the i th location will be yellow; the tile in the i th location will be blue; the next tile, the one in the $(i+1)$ th location, will be magenta and the rest of the tiles will be green. The purpose of the yellow and magenta tiles is to compute the diagonal line, marking the i th position on the i th row. Fig. 8 (its right part) shows one sample executions of the subtracting system. 2. Identity. It describes a system that ignores the input on the rightmost column, and simply copies upwards the input from the bottom row. This is a fairly straightforward system that will not need much explanation. Fig. 2 shows the 4 tiles of T_x . 3. Nondeterministic guess. It describes a system that nondeterministically decides whether or not the next w_i should be subtracted from b , Fig. 3 shows the 20 tiles of T_g . 4. Deciding Subsystem. Intuitively, we plan to write out the elements of \bar{w} on a column and b on a row, and the system will nondeterministically choose some of the elements from \bar{w} to subtract from b . Then the system will check, to make sure that no subtracted element was larger than the number it was being subtracted from, and whether $b - \sum_{i=1}^n a_i w_i \geq 0$. If it is, then a special identifier tile will attach to signify that this vector is a possible solution to the problem. Fig. 4 shows the 9 tiles of T_v and Fig. 5 shows 7 tiles of T_L .

3.4 L-Configuration Constant-Tile Adder

In Ref. [15], Brun described two addition methods, now we modify his later scheme and the adder acts on a true L-configurations to only use $\mathcal{O}(1)$ tiles to compute by invoking an idea of sandwiching tiles, this addition system that uses $\mathcal{O}(n)$ tile types to compute, but builds on L-shape seed configuration. This adder will use the two sides of the L-configuration to encode inputs, and produce its output on the top row of an almost complete rectangle. Therefore, systems could chain computations together, using the output of this computation as an input to another computation. Fig. 6 shows this system's input tiles and nondeterministic tiles.

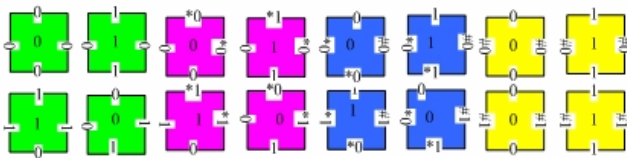


Fig. 1. There are 16 tiles in T_- . The value in the middle of each tile t represents that tile's $v(t)$ value. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

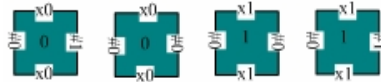


Fig. 2. There are 4 tiles in T_x . The value in the middle of each tile t represents that tile's $v(t)$ value.

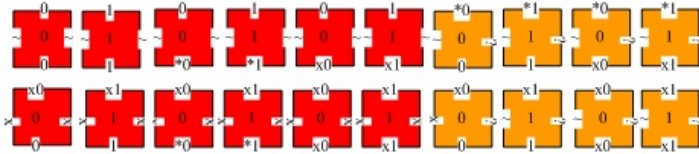


Fig. 3. There are 20 tiles in T_γ . The value in the middle of each tile t represents that tile's $v(t)$ value. Unlike the red tiles, the orange tiles do not have unique *east-south* binding domain pairs, and thus will attach nondeterministically.

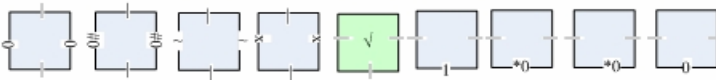


Fig. 4. There are 9 tiles in $T_\sqrt{\cdot}$. The tile with a $\sqrt{\cdot}$ in its middle will serve as the identifier tile.

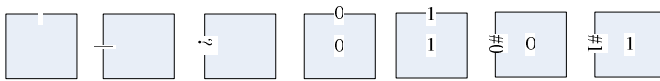


Fig. 5. There are 7 tiles in T_L . The value in the middle of each tile t represents that tile's $v(t)$ value.

In our addition system, in order to keep together with the nondeterministic tiles of the right part (see Fig. 8) and easy to calculate, we define this system's left is lower bit and right is higher bit, which shown in Fig. 8. Leftmost is a set of v_i which correspond to w_i . Fig. 7. shows the 16 tiles of T_+ . In our example, all addends and subtrahends are n bits which is the length of the largest number, if $L(v_i) < n$, the number v_i needs to be padded to be n bits long with extra 0 in its high bit, this situation also exist in subtraction system. In addition, we make the number which memory the result of addition longer than augends to ensure it can contain all the bits. In our instance, we make all addends are 6 bits while result is 8 bits (see Fig. 8). In a word, the subtraction subsystem and addition subsystem can calculate respectively except that nondeterministic step, T_γ must make the same choice: calculate or copy, in other words, we should guarantee a_i have the same value in formula (1) in the same self-assemble body. We can see clearly the sub- top row of the left part is the result of $\sum a_i v_i$ and

the top row of the right part is the result of the constraint, if $b - \sum_{i=1}^n a_i w_i \geq 0$, there will be a sign tile which has a \surd in its middle, which indicates it is a possible solution to the question.

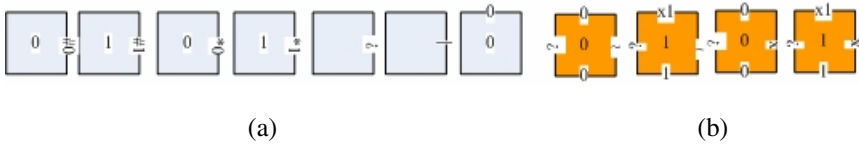


Fig. 6. (a) The input tiles of the addition system; (b) The nondeterministic tiles of the addition system

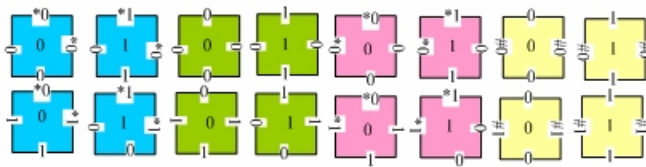


Fig. 7. Tile system S_+ computes the function $f(a, b) = a + b$ and uses 16 distinct tile types. The tiles have two input sides (*west* and *south*) and two output sides (*east* and *north*). The *north* side is the value of the bit, and the *west* side is the carry bit.

Under certain biological operations, we can obtain all the result strands which run through all the results of the 0-1 knapsack problem’s variables and the operation results tiles. Each report strand record the result of one feasible input. The strands can be amplified by polymerase chain reaction (PCR), using the primers to ligate each end of the long reporter strand. Then through gel electrophoresis and DNA sequencing, we can read out computational results. Finally, we can easily get the result of 0-1 knapsack problem. The actual implementation detail is not discussed here since they fall outside of the scope of this paper. However, we believe that we make no arbitrary hypotheses. In fact, our work is based on the assumptions and achievements that come with DNA tiling computation in general. We have described configurations that code for the correct $\vec{A} = \langle 1, 1, 0, 1 \rangle$ to allow the \surd tile to attach and $\sum a_i w_i = 01011010_2 = 90$.

Complexity analysis. The complexity of the design is considered in terms of computation time, computation space and number of distinct tiles required. It is obvious from the examples given that the computation time T is: $T = (m+2) * (n+1) * 2 = O(n)$. It follows directly from the assembly time corollary [15]. The space S taken for each assembly is the area of the assembly. $S = (n+2) * [(m * (n+1) + 1)] * 2 = O(n^2)$, which is upper-bounded polynomial to the number of variables. Finally, the library of fixed titles which type should contain the above mentioned tiles in Fig. 1 to Fig. 7: $16+4+20+9+7+16+7+4=83$.

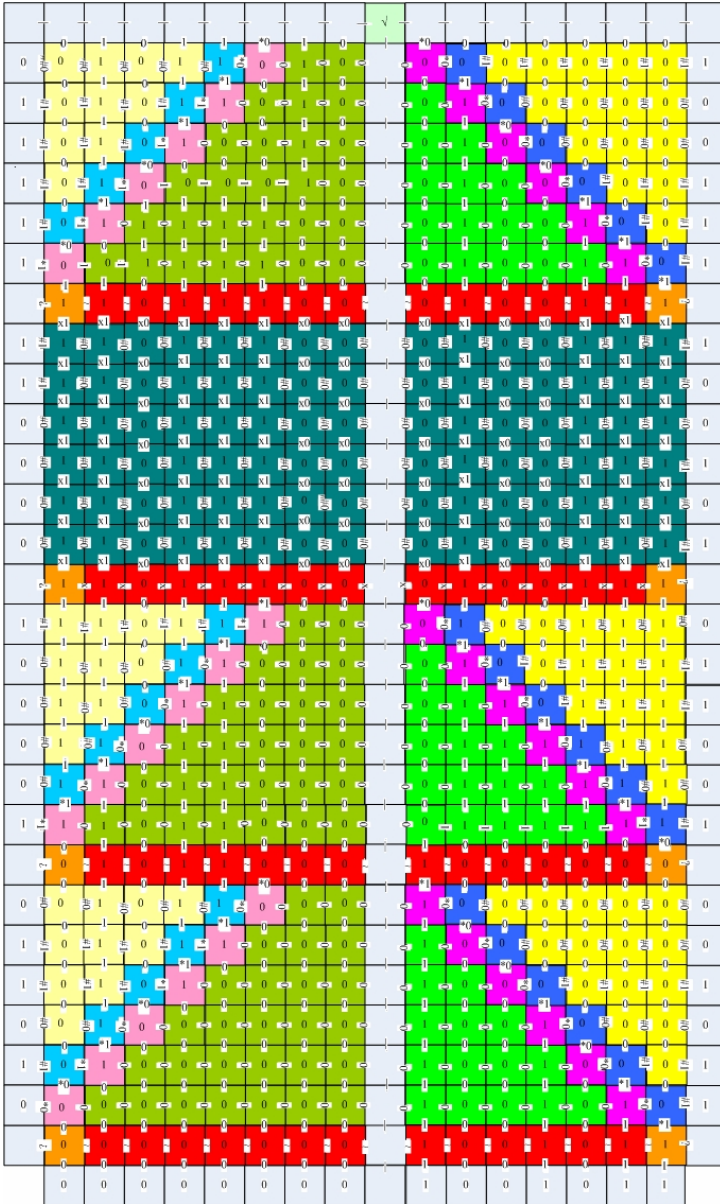


Fig. 8. An example of solving a 0-1 knapsack problem. Here, $b=75=1001011_2$, and $w_1=11=1011_2$, $w_2=25=11001_2$, $w_3=37=00101_2$, $w_4=39=100111_2$, $v_1=26=11010_2$, $v_2=33=100001_2$, $v_3=48=110000_2$, $v_4=31=011111_2$. The configuration encodes b on the bottom row and \bar{w} on the rightmost column, \bar{V} on the leftmost column. The fact that $75-(11+25+39) \geq 0$, $26+33+31=90=1011010_2$, thus the final configuration contains the $\sqrt{\quad}$ tile.

4 Conclusions

The tile assembly model is a formal model of self-assembly and crystal growth. As the massive parallelism of DNA computing, it has the unexampled dominance in solving difficult problem, especially for NP complete problem. DNA self-assembly is expected to be useful in various applications. In this study, a completely different DNA computing approach for solving optimization of 0-1 knapsack problem is derived. Here, we noticed what it means for a tile system to decide a set and explored these measures for a system that decides a NP-complete problem Subset [10] which is the security foundation of knapsack public-key cryptosystem. So we can break this kind of cryptosystem with DNA self-assembly.

This method mapping variables of the problem to DNA tiles, through encoding the stick ends of the tiles properly, molecule can hybridize in term of Watson-Crick complement that can generate all the possible solutions of the problem at the same time. The advantage of our method is that once the initial strands are constructed, each operation can compute fast parallel through the process of DNA self-assembly without any participation of manpower and the output of one computation can be directly passed as input to another computation. The time complexity of the proposed algorithm is $\mathcal{O}(n)$. The completeness and soundness of the algorithm were confirmed.

The field of DNA self-assembly holds tremendous promise. Our ultimate goal is to test the design(s) experimentally. However, there are still many technical hurdles to overcome before self-assembly can be developed into a reality. If the molecular and supramolecular word can be controlled at will, then it may be possible to achieve vastly better performance for computers and memories. We hope that this paper helps to demonstrate that molecular computing is a technology worth pursuing.

Acknowledgments. We acknowledge the support by the National Natural Science Foundation of China (Grant Nos. 60573190, 60773122).

References

1. Adleman, L.M.: Molecular Computation of Solutions to Combinatorial Problems. *Science* 266, 1021–1024 (1994)
2. Wang, H.: Proving Theorems by Pattern Recognition I. *Bell System Technical Journal* 40, 1–42 (1961)
3. Winfree, E., Eng, T., Rozenberg, G.: String Tile Models for DNA Computing by Self-Assembly. In: Condon, A., Rozenberg, G. (eds.) *DNA 2000*. LNCS, vol. 2054, pp. 63–88. Springer, Heidelberg (2001)
4. Winfree, E.: Algorithmic Self-Assembly of DNA. Ph.D. Dissertation, California Institute of Technology. Pasadena CA (1998)
5. Seeman, N.C.: DNA Nanotechnology: Novel DNA Constructions. *Annual Review of Biophysics and Biomolecular Structure* 27, 225–248 (1998)
6. Reif, J.H.: Computing: Successes and Challenges. *Science* 296, 478 (2002)
7. Rozenberg, G., Spaink, H.: DNA Computing by Blocking. *Theoretical Computer Science* 292, 653 (2003)
8. Winfree, E., Liu, F., Wenzler, L.A., Seeman, N.C.: Design and Self-Assembly of Two-Dimensional DNA Crystals. *Nature* 394, 539 (1998)

9. Mao, C., Sun, W., Seeman, N.C.: Designed Two-Dimensional DNA Holliday Junction Arrays Visualized by Atomic Force Microscopy. *J. Am. Chem. Soc.* 121, 5437 (1999)
10. Mao, C., LaBean, T.H., Reif, J.H., Seeman, N.C.: Logical Computation Using Algorithmic Self-Assembly of DNA Triple-Crossover Molecules. *Nature* 407, 493–496 (2000)
11. Robinson, R.M.: Undecidability and Nonperiodicity for Tilings of the Plane. *Inventiones Mathematicae* 3, 177 (1971)
12. Barish, R., Rothmund, P., Winfree, E.: Two Computational Primitives for Algorithmic Self-Assembly: Copying and Counting. *Nano Letters* 5(12), 2586–2592 (2005)
13. Cook, M., Rothmund, P., Winfree, E.: Self-Assembled Circuit Patterns. In: *Proceedings of the 10th International Meeting on DNA Based Computers*, pp. 91–107 (2004)
14. Rothmund, P., Papadakis, N., Winfree, E.: Algorithmic Self-Assembly of DNA Sierpinski Triangles. *PLoS Biology* 12, 2041 (2004)
15. Brun, Y.: Arithmetic Computation in the Tile Assembly Model: Addition and Multiplication. *Theoretical Computer Science* 378, 17–31 (2006)
16. Brun, Y.: Nondeterministic Polynomial Time Factoring in the Tile Assembly Model. *Theoretical Computer Science* 395, 3–23 (2008)
17. Brun, Y.: Solving NP-Complete Problems in the Tile Assembly Model. *Theoretical Computer Science* 395, 31–44 (2008)
18. Zhang, X.C., Wang, Y.F., Chen, Z.H., Xu, J., Cui, G.Z.: Arithmetic Computation Using Self-Assembly of DNA Tiles: Subtraction and Division. *Progress in Natural Science* (in print, 2009)
19. Li, X., Yang, X., Qi, J., Seeman, N.C.: Antiparallel DNA Double Crossover Molecules as Components for Nanoconstruction. *J. Am. Chem. Soc.* 118, 6131 (1996)
20. Liu, D., Park, S.H., Reif, J.H., LaBean, T.H.: DNA Nanotubes Self-Assembled from Triple-Crossover Tiles as Templates for Conductive Nanowires. *Proceedings of the National Academy of Science*, 717–722 (2004)
21. Carbone, A., Seeman, N.C.: Molecular Tiling and DNA Self-Assembly. In: Jonoska, N., Păun, G., Rozenberg, G. (eds.) *Aspects of Molecular Computing*. LNCS, vol. 2950, pp. 61–83. Springer, Heidelberg (2003)
22. Reif, J.H., LaBean, T.H., Seeman, N.C.: Challenges and Applications for Self-Assembled DNA Nanostructures. In: Condon, A., Rozenberg, G. (eds.) *DNA 2000*. LNCS, vol. 2054, pp. 173–198. Springer, Heidelberg (2001)

Learning Kernel Matrix from Gene Ontology and Annotation Data for Protein Function Prediction

Yiming Chen¹, Zhoujun Li², and Junwan Liu¹

¹ Computer School of National University of Defence and Technology,
Changsha, 410001, China

{nudtchenym,ljwnudt}@163.com

² Computer School of Beihang University,
Beijing, 100000, China
lizj@buaa.edu.cn

Abstract. During the last few years, Kernel methods have gained considerable attention for analyzing biological data for protein function prediction. Based on biological processes annotation of Yeast and GO (gene ontology), we constructed a kernel matrix to predict protein functions. We used measurement method about semantic similarity on GO and adaptive Hausdorff distance to successfully obtain protein similarity matrix, and furthermore, transformed protein similarity matrix to a undirected graph. Then, We developed a novel method that can learn optimal diffusion kernel from graph by maximizing kernel-target alignment. Experimental results illustrate that the kernel matrix generated by our formula has larger AUC value than ordinary diffusion kernel and those proposed before. Our method can even learn a common optimal kernel matrix for multiple predict tasks at one run. Furthermore, it can also be directly used to learn from various biological networks.

Keywords: Gene ontology, Diffusion kernel, Protein function prediction.

1 Introduction

Recently, predicting protein function has been becoming popular by using computational methods [1,2]. For single function class, it can be formulated as a binary classification problem. Thus, multiple function classes prediction can be transformed to a multi-class classification task.

As a popular and effective classification technology, SVM (support vector machine) is suitable for protein function prediction. Generally, this method can be divided into two steps: constructing kernel function or kernel matrix and learning SVM model. In practice, kernel matrix, which can capture essential relationship between input elements, has important influence on classification performance of SVM [3]. Selecting a suitable kernel matrix is a key step of learning a good classification model.

Many kernel matrices are defined for predicting protein function. In the paper [2], for Yeast protein datasets, Lanckriet et.al calculated eight kernel matrices based on protein sequence, expression profiles and protein-protein interaction respectively. 3,588 known proteins were labelled for 13 function categories.

GO (Gene Ontology) [4] is controlled structure vocabularies from biological research and forms a DAG (Directed Acyclic Graph) according to *is_a* and *part_of* relationship between two terms. It contains molecular function ontology, biological process ontology and cellular component ontology. A GO annotation represents a link between a gene product type and a molecular function, biological process, or cellular component type. In this work, we develop a novel kernel matrix for protein function prediction based on the idea that two proteins may have similar function if they have similar biological process or cellular component annotation.

2 Optimized Diffusion Kernel Based on GO and Annotation Data

In this section, an optimized diffusion kernel is constructed in four steps: 1, GO term similarity matrix is computed using GO DAG and biological process annotation. 2, protein similarity matrix is generated based on biological process annotation and GO term similarity matrix. 3, a undirected graph is formed by linking proteins based on protein similarity. 4, an optimized diffusion kernel matrix is obtained by using our creative method.

2.1 The Semantic Similarity between GO Terms

On GO, several measurement methods about term similarity have been proposed [5,6], they all rely on GO structure and GO term probability.

Definition 1. *The frequency of GO term is defined as number occurring in the annotation database and given by*

$$freq(c) = anno(c) + \sum_{h \in children(c)} freq(h) \quad (1)$$

anno(c) is the number of gene products annotated with this term, children(c) is the set of child nodes of term c.

Definition 2. *The probability of term c is defined as*

$$p(c) = freq(c) / freq(root) \quad (2)$$

where freq(root) is the frequency of the root term.

we prefer to use the Lin's similarity measurement. This definition equation is given by

$$Sim_{Lin}(c_1, c_2) = \max_{c \in S(c_1, c_2)} \left(\frac{2 \log p(c)}{\log p(c_1) + \log p(c_2)} \right) \quad (3)$$

$S(c_1, c_2)$ is the set of common ancestors of terms c_1 and c_2 . Obviously, the similarity value are symmetric about terms c_1 and c_2 and assuming n GO terms, we measure the similarity between them and can obtain a $n \times n$ symmetric square matrix $T = (t_{ij})_{n \times n}$.

2.2 Protein Similarity Matrix

Considering two gene products A and B , their annotations about the same GO are GO^A and GO^B , where GO^A and GO^B are term sets with $|GO^A| = m$ and $|GO^B| = n$ term elements respectively. Thus, a $m \times n$ matrix $S = (s_{ij})_{m \times n}$ is generated in which rows represent elements in GO^A , columns represent elements in GO^B and $s_{ij} = t_{l_i, k_j}$ where l_i is number of the i -th element of GO^A in GO term set and k_j is number of j -th element of GO^B in GO term set.

We use variant of Hausdorff Distance to measure the similarity between two proteins. The distances from A to B and from B to A is defined as

$$D_{hausdorff}^{A \rightarrow B} = \frac{1}{m} \sum_{i=1}^m \max_{1 \leq j \leq n} s_{ij}$$

$$D_{hausdorff}^{B \rightarrow A} = \frac{1}{n} \sum_{j=1}^n \max_{1 \leq i \leq m} s_{ij} \tag{4}$$

As the Hausdorff Distance is not symmetrical, a symmetrical measure was formulated as:

$$D_{hausdorff}^{A \leftrightarrow B} = \max\{D_{hausdorff}^{A \rightarrow B}, D_{hausdorff}^{B \rightarrow A}\} \tag{5}$$

Thus, for n proteins, we obtain a matrix $P = (p_{ij})_{n \times n}$, where $p_{ij} = D_{hausdorff}^{i \leftrightarrow j}$. It is real and symmetric matrix and called protein similarity matrix.

2.3 Diffusion Kernel of Protein Similarity Graph

According to Mercer theorem [3], kernel function and result kernel matrix should be symmetric and semi-positive definite. Although above matrix P is symmetric, it is not necessary to be semi-positive definition. To derive a kernel matrix, we transform the protein similarity matrix to a graph that is called protein similarity graph.

Definition 3. Given a similarity matrix M for dataset T , a similarity graph for a dataset T is a graph $G = (S, E)$, whose vertices are the data items in T , while the undirected links between them are labeled with some similarity measure $\tau(x, z) = M_{xz}$.

When transforming protein similarity matrix to protein similarity graph, we used a threshold value parameter γ to decide whether a link exists between two proteins. Two proteins are linked when similarity measure is larger than γ , otherwise not. In our experiments, we made $\gamma = 0.5$.

The diffusion kernel is a typical kernel defined on graph and can be described as [7]:

$$K = \lim_{s \rightarrow \infty} (I + \frac{\beta L}{s})^s = e^{\beta L} \tag{6}$$

where β is a parameter for controlling the extent of diffusion and $L \in R^{n \times n}$ is the graph Laplacian matrix defined as $L = \text{diag}(Ae) - A$ where A is the adjacency matrix, e is the vector of all ones. It is obvious that for any symmetric matrix L , $e^{\beta L}$ is always positive definite and then can be used as kernel matrix.

However, selecting an optimal parameter β becomes a challenging task. If β is too small, the local information can not be diffused effectively, resulting in a kernel matrix that only captures local similarity. On the other hand, if it is too large, the neighborhood information will be lost. Furthermore, the optimal value of β is data-dependent and it is highly desirable to tune the β value adaptively from the data.

2.4 Optimized Diffusion kernel of Protein Similarity Graph

Definition 4. The alignment $A(K_1, K_2)$ between two kernel matrices K_1 and K_2 is given by

$$A(K_1, K_2) = \frac{\langle K_1, K_2 \rangle_F}{\sqrt{\langle K_1, K_1 \rangle_F \langle K_2, K_2 \rangle_F}} \tag{7}$$

where $\langle K_1, K_2 \rangle_F$ is Frobenius inner production

$$\langle K_1, K_2 \rangle_F = K_1 \cdot K_2 = \sum_{i,j=1}^l K_{1ij} K_{2ij} = \text{trace}(K_1^T K_2)$$

[8] shows that if the value of alignment between kernel matrix and target matrix yy' is high, then there exist prediction functions that generalize well, y is a label column vector. We choose a sequence of values for β as $\beta_1 \dots \beta_p$ and learn an optimal kernel as a linear combination of corresponding diffusion kernels, requiring that the sum of combinational coefficients is equal to 1, which is motivated from the work in [9]. Thus, the optimal kernel matrix can be represented as

$$K_{opt} = \sum_{i=1}^p \alpha_i \frac{e^{\beta_i L}}{\text{trace}(e^{\beta_i L})} \quad \alpha_i \geq 0, \sum_{i=1}^p \alpha_i = 1. \tag{8}$$

Recall that the graph Laplacian matrix L is symmetric, so its eigen-decomposition can be expressed as

$$L = PDP^T. \tag{9}$$

where $D = \text{diag}(d_1, d_2, \dots, d_n)$ is the diagonal matrix of eigenvalues and $P \in R^{n \times n}$ is the orthogonal matrix of corresponding eigenvectors. According to properties of matrix function and the formula (9), we have

$$e^{\beta_i L} = PD_i P^T \tag{10}$$

where $D_i = \text{diag}(e^{\beta_i d_1}, \dots, e^{\beta_i d_n})$. and

$$\text{trace}(e^{\beta_i L}) = \text{trace}(D_i) \tag{11}$$

For single function prediction task, following theorem illustrates that optimal parameter α can be obtained by solving convex optimization problem.

Theorem 1. *For n training examples and single classification task, given p diffusion kernel $K_i = e^{\beta_i L}, i = 1, \dots, p$. and a label column vector \mathbf{y} , the combination coefficients of optimal diffusion kernel can be obtained by solving following optimization problem:*

$$\begin{aligned} & \max_{\alpha} \quad \frac{\mathbf{c}\alpha}{\alpha^T B \alpha} \\ \text{s.t.} \quad & \sum_{i=1}^p \alpha_i = 1, \alpha \geq 0 \end{aligned} \tag{12}$$

where $\alpha = (\alpha_1, \dots, \alpha_p)^T$, $c_i = (\mathbf{y}^T P) \frac{D_i}{\text{trace}(D_i)} (\mathbf{y}^T P)^T$, and $B_{ij} = \frac{\text{trace}(D_i D_j)}{\text{trace}(D_i) \text{trace}(D_j)}$.

Proof. According to (8),(10) and (11), we have:

$$\langle K_{opt}, \mathbf{y}\mathbf{y}^T \rangle_F = \sum_{i=1}^p \mathbf{y}^T K_{opt} \mathbf{y} = \sum_{i=1}^p \alpha_i (\mathbf{y}^T P) \frac{D_i}{\text{trace}(D_i)} (\mathbf{y}^T P)^T = \sum_{i=1}^p \alpha_i c_i = \mathbf{c}\alpha.$$

where $c_i = (\mathbf{y}^T P) \frac{D_i}{\text{trace}(D_i)} (\mathbf{y}^T P)^T$, $\mathbf{c} = (c_1, c_2, \dots, c_p)$

$$\langle K_{opt}, K_{opt} \rangle_F = \sum_{i,j=1}^p \alpha_i \alpha_j \frac{\text{trace}(P D_j P^T P D_i P^T)}{\text{trace}(D_j) \text{trace}(D_i)} = \sum_{i,j=1}^p \alpha_i \alpha_j B_{ij} = \alpha^T B \alpha.$$

where $B_{ij} = \frac{\text{trace}(D_i D_j)}{\text{trace}(D_i) \text{trace}(D_j)}$

According to Eq(7) and $\langle \mathbf{y}\mathbf{y}^T, \mathbf{y}\mathbf{y}^T \rangle = n^2$, we have:

$$\max_{\alpha} \quad \frac{\mathbf{c}\alpha}{n\sqrt{\alpha^T B \alpha}}$$

Which is equivalent to:

$$\max_{\alpha} \quad \frac{\mathbf{c}\alpha}{\alpha^T B \alpha}$$

For multiple function prediction tasks, we think that the graph Laplacian matrices are the same for all tasks. By assuming all tasks to share a common linear combination of kernel, we can easily obtain following joint optimization problem:

Theorem 2. *For n training examples and t classification tasks, given p diffusion kernels $K_i = e^{\beta_i L}, i = 1, \dots, p$. and t label column vectors $\mathbf{y}_k \quad k = 1, 2, \dots, t$, the combination coefficients of optimal diffusion kernel can be obtained by solving following optimization problem:*

$$\max_{\alpha} \quad \frac{\sum_{k=1}^t \mathbf{c}_k \alpha}{\alpha^T B \alpha}$$

$$\begin{aligned}
 s.t. \quad & \sum_{i=1}^p \alpha_i = 1 \\
 & \alpha \geq 0
 \end{aligned} \tag{13}$$

where $\alpha = (\alpha_1, \dots, \alpha_p)$, $c_{ki} = (\mathbf{y}_k^T P) \frac{D_i}{\text{trace}(D_i)} (\mathbf{y}_k^T P)^T$, and $B_{ij} = \frac{\text{trace}(D_i D_j)}{\text{trace}(D_i) \text{trace}(D_j)}$.

Optimization problem (12) and (13) are typical convex optimization problems [10]. The unique solution can be obtained by using some mature solving technologies.

3 Experiments

Each kernel matrix's performance is measured by performing 10-fold cross-validation using SVM classifier. For a given split, we evaluate kernel matrices by drawing the ROC(receiver operating characteristic) curve, or computing the AUC(area under curve) that is the area under ROC curve. All of the optimization formulations proposed in this paper are solved using the Matlab function *fmincon* which employs the sequential quadratic programming method [11]. SVM Classifier is trained and tested using SHOGUN machine learning toolkit [12]. All of the experiments are performed on a PC with Intel(R) Pentium(R) M processor 1.60GHz and 2G RAM.

3.1 Data Sources

We download the gene ontology file formatted as OBO1.2(Open Biological Object) and corresponding Yeast annotation file released in June 2008. For biological process, genes annotation involve 1,685 biological processes and 6,340 genes or gene products. All of the biological processes terms form a directed acyclic graph according to *is_a* or *part_of* relationship. Meanwhile, We used Lanckriet et.al's data, it contains 6,351 genes or gene products in which 3,588 genes or genes products is known and 13 function classification tasks [2].

3.2 The ROC Comparison of Optimal Diffusion Kernel with Ordinary Diffusion Kernel

To compare the ROC performance of optimal diffusion kernel with other diffusion kernels, we compute five ordinary diffusion kernels when diffusion constants $\beta = -0.1, -0.5, -1, -1.5, -2$ respectively and two composite diffusion kernels, one is linear combinational of 20 diffusion kernels with same weights, the other is optimal diffusion kernel on 20 diffusion kernels generated by our formula. Four function classes, these are metabolism, energy, cell cycle and transcription, are randomly selected for plotting ROC curves by using 10-fold cross-validation. These ROC curves are shown in Fig.1.

It can be observed that the ROC performance of the optimal diffusion kernel is consistently better than that of other diffusion kernels, although composite diffusion kernel with same weight 0.05 is better than ordinary those in most cases.

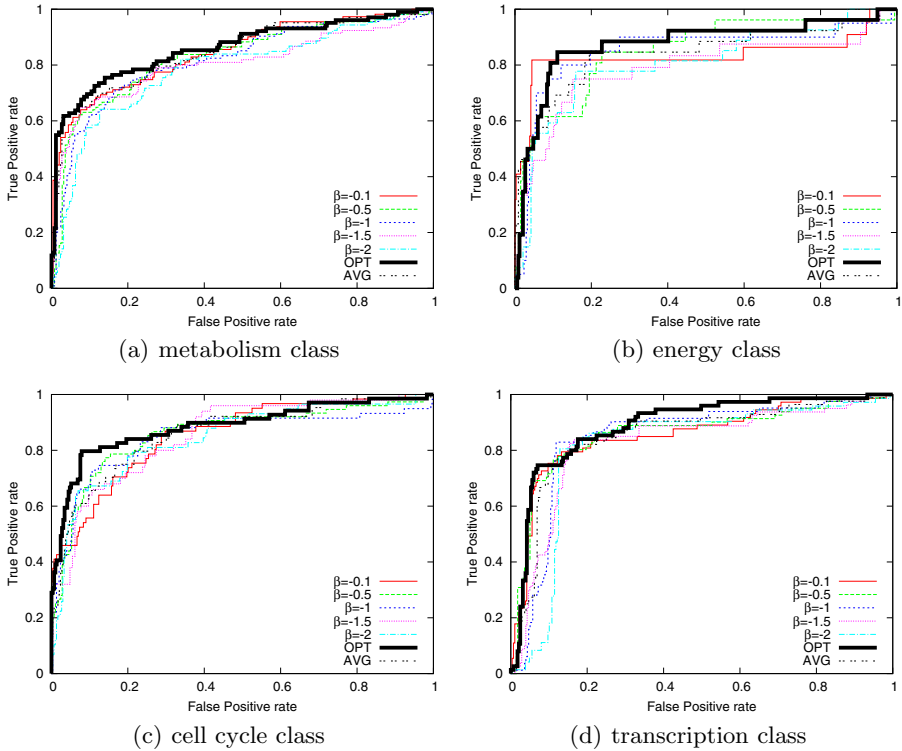


Fig. 1. Comparing the ROCs about five ordinary diffusion kernels and two composite diffusion kernel for four function classes. One is composite with same weight, the other is optimal diffusion kernel. The ROC performance of the optimal diffusion kernels are better than other kernels.

3.3 The ROC Comparison of Optimal Diffusion Kernel with Lanckriet’s Kernels

We select four kernel matrices that are representative and have best ROC performance in homogeneous kernels out of [2]. They are KEG(kernel matrix of expression data with Gaussian kernel), KPI(kernel matrix of physical protein-protein interactions data with diffusion kernel), PFD(kernel matrix of protein structure domain sequence data with inner product kernel) and KSW(kernel matrix of sequence data with Smith Waterman alignment). Similarly, we compare the ROC performance of optimal diffusion kernel generated by our formula with selected those above on four function classes same as the section 3.3. The results are shown in Fig.2.

For selected four function classes, KSW has the best ROC performance in Lanckriet’s kernels. However, the ROC performance of the OPT is better than those of Lanckriet’s kernel matrices though, for metabolism class, the ROC

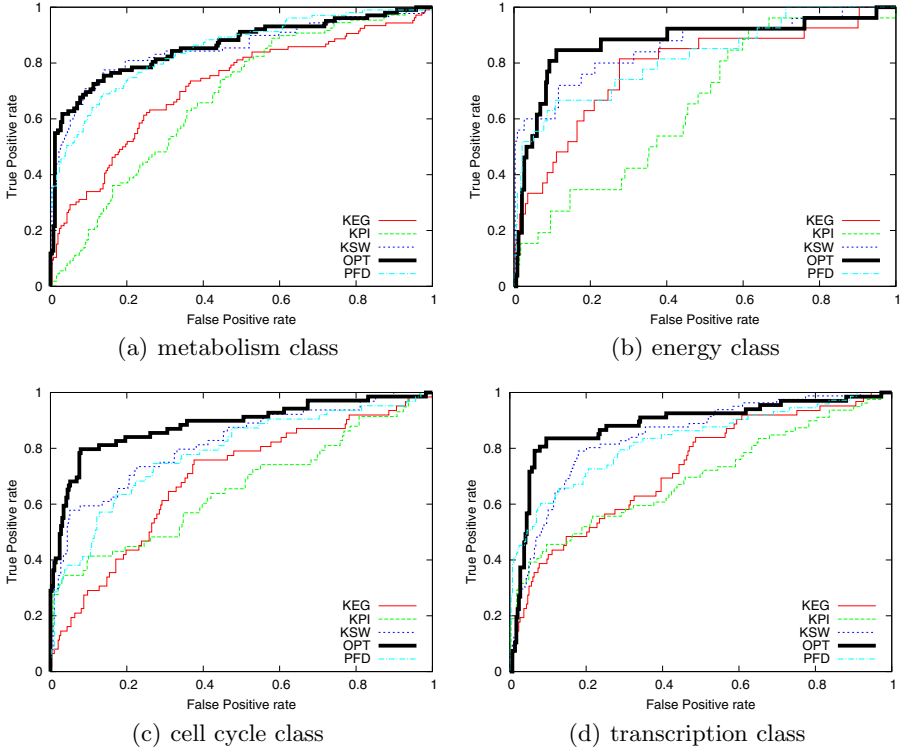


Fig. 2. The ROC performance comparison of OPT(optimal diffusion kernel) with Lanckriet’s kernels, that is KEG(kernel matrix of expression data with Gaussian kernel), KPI(kernel matrix of physical protein-protein interactions data with diffusion kernel), PFD(kernel matrix of protein structure domain sequence data with inner product kernel) and KSW(kernel matrix of sequence data with Smith Waterman alignment)

performance of the OPT is little better than that of KSW. The AUC of the OPT is 0.8601 while that of the KSW is 0.8528.

3.4 The Optimal Diffusion Kernel in Multi-task Case

Theorem 2 indicates that we can obtain a common optimal diffusion kernel for 13 function classes by solving a convex programming problem, which is called multi-task learning. We compute the common optimal diffusion kernel and 13 optimal diffusion kernels for single function class. In Fig.3., the bar graph shows AUCs pair for 13 function classes.

Fig.3. shows that, in multi-task learning case, we can learn a common optimal diffusion kernel at one run because its AUC is only little less than that of single task learning. In some cases, its AUC is even larger than that of single task learning.

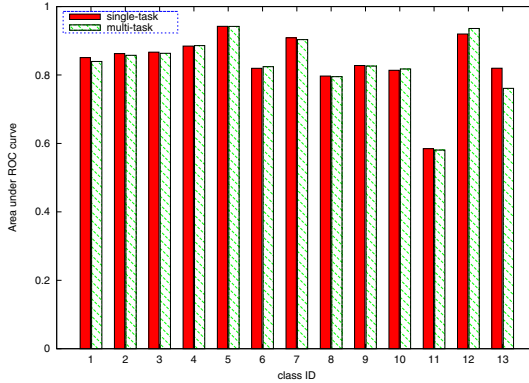


Fig. 3. The AUC comparison of diffusion kernels between multi-task and single task learning for 13 function classes

4 Conclusions

We learn a kernel matrix from biological process ontology and corresponding annotation data. The experimental results show that the kernel matrix generated by our formulas is better than typical those. Furthermore, our method can learn a common optimal kernel matrix for several function prediction task at one run. Lastly, our method can also be directly applied to learn optimal diffusion kernel from biological networks for protein function prediction.

5 Future Works

Since it has been theoretically proven that the kernel matrix with larger kernel-target alignment value certainly has a good generalization performance when predicting [8], we do not currently validate it by predicting the function of unknown protein. It should be our next work. Using our method, another optimal kernel matrix can be computed by using cellular component ontology and corresponding annotation data. If we combine these two kernel matrices by using MKL(multiple kernel learning) proposed by Lanckriet et.al [9], more better prediction performance may be obtained.

Acknowledgments. This work is in part supported by the National Scientific Foundations in P.R. China(Grant No: 60573057) and the scientific foundations of Hunan Agricultural University, P.R. China(Grant No: 06YJ16).

References

1. Tsuda, K., Noble, W.S.: Learning Kernels from Biological Networks by Maximizing Entropy. *Bioinformatics* 20, 326–333 (2004)
2. Lanckriet, G., Deng, M., Cristianini, M., Jordan, M., Noble, W.: Kernel-based Data Fusion and Its Application to Protein Function Prediction in Yeast. In: *Pac. Symp. Biocomput.*, vol. 9, pp. 300–311 (2004)

3. Cristianini, N., Shawe-Taylor, J.: *Kernel Methods for Pattern Analysis*. China Machine Press, Beijing (2005)
4. Ashburner, M., Ball, C., Blake, J., Botstein, D.: Gene Ontology: Tool for the Unification of Biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–29 (2000)
5. Resnik, P.: Semantic Similarity in A Taxonomy: An Information-based Measure and Its Application to Problems of Ambiguity in Natural Language. *J. Artif. Intell. Res.* 11, 95–130 (1999)
6. Lin, D.: An Information-theoretic Definition of Similarity. In: *The 15th International Conference on Machine learning*, pp. 296–304. Morgan Kaufmann Publishers, San Francisco (1998)
7. Kondor, R.I., Lafferty, J.: Diffusion Kernels on Graphs and other Discrete Input Spaces. In: *Proc. Int. Conf. Machine Learning*, pp. 315–322. Morgan Kaufmann Publishers, San Francisco (2002)
8. Nello, C., John, S., Elissee, J., Kandola, A.: On Kernel-target Alignment. In: *Advances in Neural Information Processing Systems 14*, pp. 367–373. MIT Press, Cambridge (2002)
9. Lanckriet, G.R., Cristianini, N., Bartlett, P., Laurent, E., Michael, I.G., Jordan, M.I.: Learning the Kernel Matrix with Semidefinite Programming. *Journal of Machine Learning Research* 5, 27–72 (2004)
10. Boyd, S., Vandenberghe, L.: *Convex Optimization*. Cambridge University Press, London (2004)
11. The Mathworks Incorporation: Matlab help, <http://www.mathworks.cn/>
12. Sonnenburg, S., Raetsch, G., Schaefer, C., Scholkopf, B.: Large scale multiple kernel learning. *Journal of Machine Learning Research* 7, 1531–1565 (2006)

Improved Quantum Evolutionary Algorithm Combined with Chaos and Its Application^{*}

Jianhua Xiao

Research Center of Logistics, Nankai University, Tianjin 300071, China
Department of Control Science and Engineering, Huazhong University of Science and
Technology, Wuhan, 430074, China
jhxiao2008@163.com, jhxiao@nankai.edu.cn

Abstract. Quantum evolutionary algorithm (QEA) has been developed rapidly and has been applied widely during the past decade. In this paper, an improved quantum evolutionary algorithm (IQEA) is presented based on particle swarm optimization (PSO) and chaos. The simulation results in solving DNA encoding demonstrate that the improved quantum evolutionary algorithm is valid and outperforms the quantum chaotic swarm evolutionary algorithm and conventional evolutionary algorithm. *abstract* environment.

Keywords: Quantum evolutionary algorithm, Particle swarm optimization, Chaos Optimization, DNA encoding problem.

1 Introduction

Since quantum computing was proposed by Benioff and Eeynman [1] [2] in the early 1980s, it has developed rapidly and has shown significant potential in solving various difficult problems. In 1996, Narayanan and Moore [3] first introduced quantum evolutionary algorithm, which was inspired by the concept of quantum computing. Han et al. [4] gave its practical form. In recent years, quantum evolutionary computing was also used to solve various optimization hard problems. Jiang et al. [5] applied quantum evolutionary algorithm to solve face verification. Yang et al. [6] proposed a new blind source separation method based on quantum genetic algorithm, and the simulation result showed that the effect of the new method is greater than that of conventional genetic algorithm (CGA). Li et al. [7] proposed a hybrid quantum-inspired genetic algorithm for a multi-objective flow shop scheduling problem. Feng et al. [8] developed a novel quantum coding mechanism to solve the traveling salesman problem (TSP)

^{*} This work was supported by the National Natural Science Foundation of China (Grant Nos. 60674106, 60703047, and 60533010), the Program for New Century Excellent Talents in University (NCET-05-0612), the Ph.D. Programs Foundation of Ministry of Education of China (20060487014), the Chenguang Program of Wuhan (200750731262), 2008 Program Project of Humanity and Social Science of Nankai University (NKQ08058), and HUST-SRF (2007Z015A).

based on the quantum-inspired evolutionary algorithm. Wang et al. [9] designed a quantum ant colony optimization algorithm (QACO) to solve the discrete optimization. Wang et al. [10] proposed a novel quantum swarm evolutionary algorithm (QSEA) to solve the 0-1 knapsack problem.

In this paper, an improved quantum evolutionary algorithm is proposed, which is based on QEA. The proposed algorithm adopts the improved PSO combined with chaos to update the Q-bit automatically. The simulation results in solving DNA encoding demonstrate that IQEA is valid and outperforms the quantum chaotic swarm evolutionary algorithm (QCSEA) and conventional evolutionary algorithm.

The rest of the paper is organized as follows. The conventional quantum evolutionary algorithm is introduced in section 2. The improved quantum evolutionary algorithm combined with chaos is proposed in section 3. The simulation results in solving DNA encoding are shown and discussed in section 3. The conclusion and further remarks are given in section 4.

2 Quantum Evolutionary Algorithm

Quantum computing is a new computing based on the concepts and principles of quantum theory, such as superposition of quantum states, entanglement and intervention. Quantum evolutionary algorithm is inspired by concepts of quantum computing such as quantum bits and quantum gate.

In QEA, the smallest information unit is called Q-bit, which is defined by a pair of numbers (α, β) as

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} \quad (1)$$

where α and β are complex numbers that specify the probability amplitudes of the Q-bit states. The modulus $|\alpha|^2$ and $|\beta|^2$ give the probabilities that the Q-bit will be the state "0" and the state "1", respectively, which satisfy that $|\alpha|^2 + |\beta|^2 = 1$. For an m -Q-bit individual x_i , it is defined as:

$$x_i = \left[\begin{array}{c|c|c} \alpha_1 & \alpha_2 & \dots & \alpha_m \\ \beta_1 & \beta_2 & & \beta_m \end{array} \right], \quad (2)$$

where $|\alpha_i|^2 + |\beta_i|^2 = 1, (i = 1, 2, \dots, m)$.

To evaluate each individual's fitness for guiding updating of the algorithm and solve the optimization problems, the corresponding binary solution needs to be obtained by observing the state of the Q-bits. For a bit p_i of the binary individual P , a random number λ in interval $[0, 1]$ is generated and compared with $|\alpha_i|^2$ of the Q-bit individual Q . If α_i satisfies $|\alpha_i|^2 > \lambda$, then set $p_i=1$, otherwise set $p_i=0$. By these steps, whole binary solutions can be constructed by observing the states of the current Q-bit solutions. Then the fitness of each individual is evaluated by the corresponding binary solution observed. Finally, a quantum rotation gate $U(t)$ is employed to update the Q-bit individual as follows [7]:

$$\begin{bmatrix} \alpha'_i \\ \beta'_i \end{bmatrix} = \begin{bmatrix} \cos \theta_i & -\sin \theta_i \\ \sin \theta_i & \cos \theta_i \end{bmatrix} \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix}, \tag{3}$$

$$\theta_i = s(\alpha_i, \beta_i) \cdot \Delta\theta_i, \tag{4}$$

where θ_i is the rotation angle, $s(\alpha_i, \beta_i)$ and $\Delta\theta_i$ represent the sign of θ_i determining the rotation direction and the magnitude of rotation angle, respectively. The values of $s(\alpha_i, \beta_i)$ and $\Delta\theta_i$ are determined by the lookup table as shown in Table 1.

The basic pseudocode algorithm for QEA is shown in Fig. 1.

Table 1. The lookup table of rotation angle θ_i , where $f(\cdot)$ is the fitness, $s(\alpha_i, \beta_i)$ is the sign of θ_i , p_i and r_i are the i th bits of the current best solution P and the binary solution B

p_i	b_i	$f(P) > f(B)$	$\Delta\theta_i$	$S(\alpha_i, \beta_i)$			
				$\alpha_i\beta_i > 0$	$\alpha_i\beta_i < 0$	$\alpha_i = 0$	$\beta_i = 0$
0	0	False	0	0	0	0	0
0	0	True	0	0	0	0	0
0	1	False	0	0	0	0	0
0	1	True	0.01π	-1	+1	± 1	0
1	0	False	0.01π	-1	+1	± 1	0
1	0	True	0.01π	+1	-1	0	± 1
1	1	False	0.01π	+1	-1	0	± 1
1	1	True	0.01π	+1	-1	0	± 1

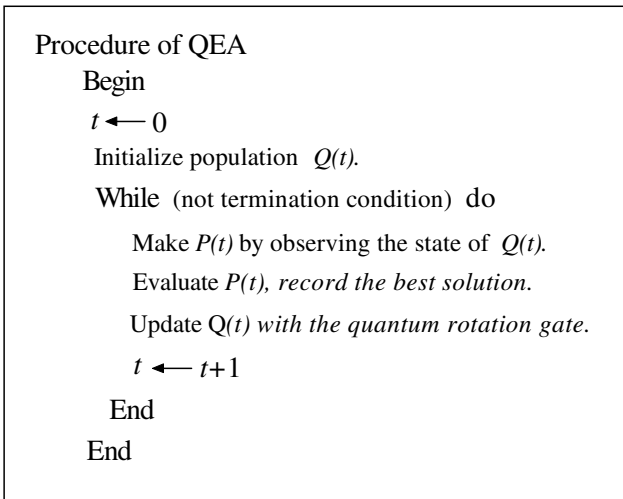


Fig. 1. Pseudocode algorithm for QEA

3 Improved Quantum Evolutionary Algorithm

In a conventional QEA, quantum rotation gate was used to update the Q-bit individuals by using lookup table. However, it had not the theoretical of the value of the rotation angle θ_i . So, Wang et al. [10] defined a new Q-bit expression, and implemented an improved particle swarm optimization to update the quantum angles automatically.

In [10], a Q-bit is presented as $[\theta]$, where $[\theta]$ is equivalent to the original Q-bit as $[\sin(\theta) \cos(\theta)]^T$. Obviously, $|\sin(\theta)|^2 + |\cos(\theta)|^2 = 1$, an m -Q-bits $\begin{bmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_m \\ \beta_1 & \beta_2 & \dots & \beta_m \end{bmatrix}$ can be replaced by $[\theta_1|\theta_2|\dots|\theta_m]$, and the corresponding rotation gate angle was replaced by $[\theta_i] = [\theta_i + \Delta\theta_i]$. Therefore, a new Q-bit expression was defined, and the particle swarm optimization (PSO) was employed to update the quantum angles automatically. The update of quantum angle is defined as follows:

$$v_{ji}^{t+1} = \omega * v_{ji}^t + C_1 * r_1 * (\theta_{ji}^t(pbest) - \theta_{ji}^t) + C_2 * r_2 * (\theta_{ji}^t(gbest) - \theta_{ji}^t) \tag{5}$$

$$\theta_{ji}^{t+1} = \theta_{ji}^t + v_{ji}^{t+1}, \tag{6}$$

where v_{ji}^t , θ_{ji}^t , $\theta_{ji}^t(pbest)$, $\theta_{ji}^t(gbest)$ are the velocity, current position, individual best and global best of i th Q-bit of the j th m -bits, respectively. ω is the inertia weight factor, C_1 and C_2 are the acceleration constants, r_1 and r_2 are two independently uniformly distributed random values in the range $[0, 1]$.

In particle swarm optimization, the parameters r_1 and r_2 are important, and can affect the PSO's convergence. As a novel optimization technique, chaos has gained much attention and some applications. Owing to the ergodicity, randomness, regularity and special ability of avoiding being trapped in local optimal solution, the chaotic search has been introduced into the various optimization algorithms, such as chaotic neural network [11], chaotic simulated annealing algorithm [12], mutative scale chaotic optimization algorithm [13], and so on. In the paper, we introduce chaotic sequences based on Bernoulli shift map [14] to improved the global convergence in substitution of parameters r_1 and r_2 .

The Bernoulli shift map belongs to class of piecewise linear maps, which consist of a number of piecewise linear segments. The Bernoulli shift map is given as follows [14]:

$$z_{i+1} = \begin{cases} \frac{z_i}{1-\lambda} & 0 < z_i \leq 1 - \lambda \\ \frac{z_i - (1-\lambda)}{\lambda} & 1 - \lambda < z_i < 1 \end{cases} \tag{7}$$

where z_i is chaotic variable. The special case of the Bernoulli shift map is

$$x_{n+1} = 2 * x_n \text{ mod } 1 \tag{8}$$

The updated equation Eq.(5) is modified as follows:

$$v_{ji}^{t+1} = \omega * v_{ji}^t + C_1 * z_{ji}^t * (\theta_{ji}^t(pbest) - \theta_{ji}^t) + C_2 * Z_{ji}^t * (\theta_{ji}^t(gbest) - \theta_{ji}^t) \tag{9}$$

where z_{ji}^t and Z_{ji}^t are given values by Bernoulli shift map between 0 and 1.

The inertia weight w is the the parameter that controls the impact of previous velocity on the current one, and can adjust the balance between exploration and exploitation. In the published literature, there are many methods to decide the inertia weight factor, such as the linear decreasing inertia weight [15], adaptive inertia weight factor (AIWF) [16], and dynamic inertia weight (DIW) [17]. In the paper, we will use the dynamic inertia weight that decreases according to iterative generation increasing. The corresponding equation is defined as follows:

$$w = w' * u^{-iter} \quad (10)$$

where $w' \in [0, 1]$, $u \in [1.0001, 1.0005]$, and $iter$ denotes the current iteration. In the paper, we will set w' to 0.72, and set u to 1.0002.

The procedure of improved quantum evolutionary algorithm is described as follows:

Step 1. Initialize population, use quantum angle to encode Q -bit.

$$Q(t) = \{q_1^t, q_2^t, \dots, q_n^t\}, \quad q_j^t = [\theta_{j1}^t | \theta_{j2}^t | \dots | \theta_{jm}^t].$$

Step 2: Make each P_{ji}^t by observing the state of $Q(t)$ through comparing with $|\cos(\theta_{ji})|^2$ or $|\sin(\theta_{ji})|^2$ as follows:

$$P_{ji}^t = \begin{cases} 0, & \text{if } rand[0, 1] > |\cos(\theta_{ji}^t)|^2 \\ 1, & \text{otherwise.} \end{cases}$$

Step 3: Calculating the fitness of population $P(t)$ by fitness function, and save the best solution.

Step 4: If stopping condition is satisfied, then output results; otherwise, go to step 5.

Step 5: Update $Q(t)$ with the following PSO formulate instead of using traditional quantum rotation angle.

$$\nu_{ji}^{t+1} = \omega * \nu_{ji}^t + C_1 * z_{ji}^t * (\theta_{ji}^t(pbest) - \theta_{ji}^t) + C_2 * Z_{ji}^t * (\theta_{ji}^t(gbest) - \theta_{ji}^t)$$

$$\theta_{ji}^{t+1} = \theta_{ji}^t + \nu_{ji}^{t+1}$$

Step 6: If stopping condition is satisfied, then output results; otherwise, let $t=t+1$, and go back to step 2.

4 Experimental Results

4.1 DNA Encoding Problem

DNA encoding problem is to design a set of DNA sequences with equal length, which satisfy some physical, chemical and logical constraints in order to avoid

mishybridization in DNA computation. In [18], the constraints consist of H-measure constraint, similarity constraint, Hairpin constraint, continuity constraint, melting temperature constraint, and GC content constraint. Each constraint function needs to be optimized simultaneously. Obviously, the optimization problem of the constraint function is the multi-objective optimization problem.

Formally, the DNA encoding problem can be written as follows:

$$\begin{aligned} \text{Optimize } & F(x) = (f_1(x), f_2(x), \dots, f_n(x)) \\ & f_i(x) \in \{ \text{fitness measures in [18]} \} \end{aligned} \tag{11}$$

The fitness function was the weighted sum of required fitness measures.

$$\begin{aligned} \text{Fitness } (x) &= \sum_i w_i f_i(x) \\ & f_i(x) \in \{ \text{fitness measures in [18]} \} \end{aligned} \tag{12}$$

For simplicity, we set each weight w_i to one. For more notation and details about DNA encoding constraint function, please refer to [18].

4.2 Results and Analyses

To illustrate the effectiveness and performance of improved quantum evolutionary algorithm for DNA encoding problem, we compare IQEA with other approaches. In simulation, the parameters is given in Table 2.

Table 2. Parameter values for DNA encoding

Quantity	Value
The population size	20
Maximum number of iterations	1000
DNA sequence length	20
The acceleration constant C_1	1.43
The acceleration constants C_2	1.58

First, we compare our algorithm with Ref. [18]. In Ref. [18], a quantum chaotic swarm evolutionary algorithm is used to select good DNA sequences for Adleman’s experimentation [20]. The set of seven DNA sequences whose length is 20-mer and corresponding fitness values are listed in Table 3. The comparison results are shown in Fig. 2.

From Fig. 2 and Table 3, it can be found that the DNA sequences generated by improved algorithm have lower H-measure values than the DNA sequences from [18], and can reduce the probability to hybridize with the noncomplementary sequences. The second structure of the DNA sequences generated by our

Table 3. Comparison results of the sequences in Ref. [18] and our sequences

DNA Sequence (5'→3')	Continuity	Hairpin	H-measure	Similarity	T _m	GC %
Our Sequences						
ATGAGCAACTTAACCGTACC	0	0	63	55	51.1693	45
AATATCCGGCGCACGTACCT	0	3	65	56	57.2020	55
CAACCAATTC AATCACCTCG	0	0	61	51	50.8097	45
CTTGCTTCTACCGACCTGTG	0	0	62	55	54.1311	55
TATCCGCCGCAGATTCTAGT	0	0	66	59	53.8833	50
TCGCGTTTCCTAGACTTCTG	9	0	58	55	53.0982	50
TGTTCCGATGTGGTCGTTTG	9	0	59	51	54.5007	50
Xiao et al.						
AACAATGAATGGGCAGGAGT	9	3	54	56	52.9306	45
CAGGACTAAACAATTCCAAA	18	3	53	60	46.9346	35
CACATTACGCCAAGGATACC	0	0	54	53	52.2051	50
GACCGCAAGACAGAAGAGAA	0	0	48	61	53.3654	50
ACCGACGTCCGTAAC TGACC	0	0	59	54	57.7230	60
ACATGAGATCAACCTGCGCA	0	0	54	56	55.6584	50
TAAGAGAATGCCAGAATAAG	0	0	50	60	45.5851	35

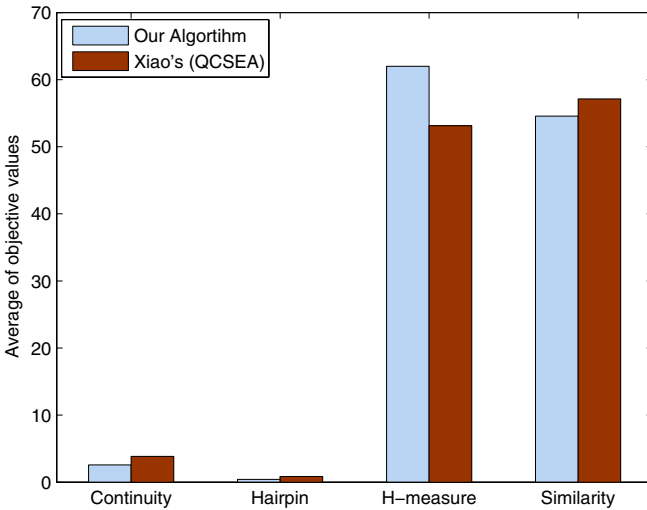


Fig. 2. Average objective values comparison between QCSEA and our algorithm

algorithm is more restrained because of the low *continuity* and *hairpin*. Furthermore, the range of melting temperatures (from 50.8097 to 57.2020) is better than QCSEA algorithm (from 45.5851 to 57.7230), which has more advantage in keeping an uniform melting temperature.

Then, our algorithm is compared with the conventional evolutionary algorithm (CEA). In [20], Shin et al. gives DNA sequences for the TSP using conventional

Table 4. Comparison results of the sequences in Ref. [20] and our sequences

DNA Sequence (5'→3')	Conti- nuity	Hairpin	H-measure	Simil- arity	T _m	GC %
Our Sequences						
GAGTCCATATAGCATCCGCC	0	0	66	55	53.3870	55
TTAATAACCGACCAGCGGAA	0	0	62	53	52.3872	45
CTCATACGCTTTGGTAGACA	9	3	61	53	50.5641	45
CGAACGGTAGCTTATAGGAA	0	0	61	58	50.0823	45
GAAGGCCACGCTACACGCAG	0	0	62	50	59.7662	65
AATGAAGAAGCGAAGACGTA	0	0	55	54	50.4292	40
TGACTGTTGCCTATGTGCGGT	0	6	65	45	54.3720	50
Shin et al.						
AGGCGAGTATGGGGTATATC	16	0	66	48	47.6070	50
CCTGTCAACATTGACGCTCA	0	3	66	57	50.6204	50
TTATGATTCCACTGGCGCTC	0	0	61	58	50.1205	50
ATCGTACTCATGGTCCCTAC	9	0	64	54	47.8464	50
CGCTCCATCCTTGATCGTTT	9	0	62	58	50.4628	50
CTTCGCTGCTGATAACCTCA	0	3	68	54	49.8103	50
GAGTTAGATGTCACGTCACG	0	3	67	51	48.3995	50

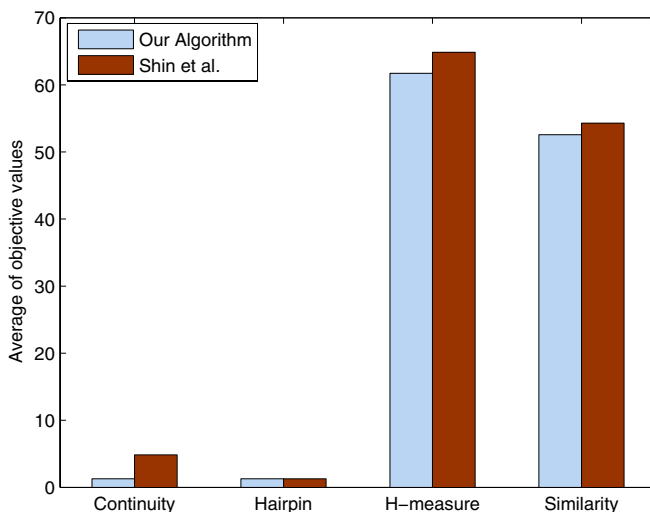


Fig. 3. Average objective values comparison between CEA and our algorithm

evolutionary algorithm (CEA). The corresponding seven DNA sequences are listed in table [4]. The comparison results in term of average of fitness are shown in Fig. [3]. From Table [4] and Fig. [3], it is clear that the proposed algorithm performed better than the conventional evolutionary algorithm according to the average of fitness values (*H-measure*, *continuity*, *similarity*). The two methods performed the same according to *hairpin* fitness.

5 Conclusions

In this paper, an improved quantum evolutionary algorithm combined with chaos is proposed. By introducing the particle swarm algorithm with DIW and chaos, the improved algorithm can avoid the disadvantage of easily getting into the local optimal solution in the later evolution period. Furthermore, the results in solving DNA encoding show that the proposed algorithm is efficient to find a set of DNA sequences with good quality. In future work, we will investigate how our algorithm can further generate much larger sets and be applied to solve other optimization problems.

References

1. Benioff, P.: The Computer as A Physical System: A Microscopic Quantum Mechanical Hamiltonian Model of Computers as Represented by Turing Machines. *J. Stat. Phys.* 22, 563–591 (1980)
2. Feynman, R.: Simulating Physics with Computers. *Int. J. Theoret. Phys.* 21, 467–488 (1982)
3. Narayanan, A., Moore, M.: Quantum-Inspired Genetic Algorithm. In: *Proceedings of IEEE International Conference on Evolutionary Computation*, pp. 61–66. IEEE Press, Nagoya (1996)
4. Han, K.H., Kim, J.H.: Genetic Quantum Algorithm and Its Application to Combinatorial Optimization Problem. In: *Proceedings of the 2000 IEEE Congress on Evolutionary Computation*, pp. 1354–1360. IEEE Press, San Diego (2000)
5. Jiang, J.S., Han, K.H., Kim, J.H.: Quantum-Inspired Evolutionary Algorithm-Based Face Verification. In: Cantu-Paz, E., Davis, L.D., Deb, K., Roy, R., Foster, J.A. (eds.) *GECCO 2003*. LNCS, vol. 2724, pp. 2147–2156. Springer, Heidelberg (2003)
6. Yang, J.A., Li, B., Zhuang, Z.Q., Zhong, Z.F.: Quantum Genetic Algorithm and Its Application Research in Blind Source Separation. *Mini-Micro System* 24, 1518–1523 (2003)
7. Li, B.B., Wang, L.: A Hybrid Quantum-Inspired Genetic Algorithm for Multi-Objective Scheduling. In: Huang, D.-S., Li, K., Irwin, G.W. (eds.) *ICIC 2006*. LNCS, vol. 4113, pp. 511–522. Springer, Heidelberg (2006)
8. Feng, X.Y., Wang, Y., Ge, H.W., et al.: Quantum-Inspired Evolutionary Algorithm for Travelling Salesman Problem. In: Bredenfeld, A., Jacoff, A., Noda, I., Takahashi, Y. (eds.) *RoboCup 2005*. LNCS, vol. 4020, pp. 1363–1367. Springer, Heidelberg (2006)
9. Wang, L., Liu, Q., Fei, M.R.: A Novel Quantum Ant Colony Optimization Algorithm. In: Li, K., Fei, M., Irwin, G.W., Ma, S. (eds.) *LSMS 2007*. LNCS, vol. 4688, pp. 277–286. Springer, Heidelberg (2007)
10. Wang, Y., Feng, X.Y., Huang, Y.X., et al.: A Novel Quantum Swarm Evolutionary Algorithm and Its Applications. *Neurocomputing* 70, 633–640 (2007)
11. Aihara, K., Takabe, T., Toyoda, M.: Chaotic Neural Network. *Physics Letter A* 144, 333–340 (1990)
12. Wang, Z., Zhang, T., Wang, H.: Simulated Annealing Algorithm of Optimization Based on Chaotic Variable. *Control and Decision* 14, 381–384 (1998)
13. Zhang, T., Wang, H., Wang, Z.: Mutative Scale Chaos Optimization Algorithm and Its Application. *Control and Decision* 14, 285–288 (1999)

14. Tavazoei, M.S., Haeri, M.: Comparison of Different One-Dimensional Maps as Chaotic Search Pattern in Chaos Optimization Algorithm. *Application Mathematics and Computation* 187, 1076–1085 (2007)
15. Shim, Y.H., Kennedy, J.: Empirical Study of Particle Swarm Optimization. In: *Proceedings of Congress on Evolutionary Computation*, Piscataway, NJ, pp. 1945–1950 (1999)
16. Liu, B., Wang, L., Jin, Y.H., et al.: Improved Particle Swarm Optimization Combined with Chaos. *Chaos, Solitons & Fractals* 25, 1261–1271 (2005)
17. Jiao, B., Lian, Z.G., Gu, X.S.: A Dynamic Inertia Weight Particle Swarm Optimization Algorithm. *Chaos, Solitons & Fractals* 37, 698–705 (2008)
18. Xiao, J., Xu, J., Chen, Z., et al.: A Hybrid Quantum Chaotic Swarm Evolutionary Algorithm for DNA Encoding. *Computers and Mathematics with Applications* (2008) doi: 10.1016/j.camwa
19. Adleman, L.M.: Molecular Computation of Solutions to Combinatorial Problems. *Science* 266, 1021–1024 (1994)
20. Shin, S.Y., Lee, I.H., Kim, D., et al.: Multi-Objective Evolutionary Optimization of DNA Sequences for Reliable DNA Computing. *IEEE Transactions on Evolutionary Computation* 9, 143–158 (2005)

Fault Diagnosis of Nonlinear Analog Circuits Using Neural Networks and Multi-Space Transformations

Yigang He and Wenji Zhu

College of Electrical and Information Engineering, Hunan University,
Hunan, Changsha 410082, China
hyghnu@yahoo.com.cn

Abstract. A systematic approach, based on piece-wise linear (PWL) models, a bilinear transformation in multidimensional spaces and back-propagation neural networks (BPNN), for nonlinear analog fault diagnosis is proposed in this paper. The functions of input-output are applied for fault diagnosis to deal with the circuits without sufficient accessible nodes. Besides, we used the functions transformation in multi-space to select fault features, which can decrease the ambiguity groups and improve the performance of fault diagnosis. Through preprocessing of the signals of test nodes from the analog circuits, the optimal features are selected. These features are then fed into the BPNNs for fault location. The single fault diagnosis is mainly discussed in this paper. Finally, an illustration to demonstrate the strength of our proposed method is given.

Keywords: Analog Circuits, Neural Network, Fault Diagnosis, Bilinear Transformation, Space Transformation.

1 Introduction

Analog fault diagnosis has been an active area of research since the mid-1970s with the significant work carried out at the system, board, and chip level, and many works appeared in the literatures [1-3]. A survey of the research conducted in this area clearly indicates that analog fault diagnosis is complicated due to the poor fault models, component tolerances, and nonlinearity issues [1-2]. As a consequent, fault diagnosis techniques for analog circuits and systems are more complex when compared to their counterparts in digital circuits [1-3]. Since, sometimes, measurements access to internal nodes of a circuit under test (CUT) in practice is difficult, new input-out testing and diagnostic methods are needed. Some possibilities for synthesis and development of such methods are offered by space transformations.

Applications of space transformations to analog fault diagnosis are discussed in [2-5]. The work in reference [4] is a pioneering effort to take bilinear transformation for parametric (soft) fault localization and identification via measurement of the real and the imaginary part of the circuit function. The faulty components can be localized by the location of the measurement point on a particular locus, whose scale gives the possibility of fault identification. However, it has been difficult to implement the method in practice, because in many cases parameter loci are situated too close to each other or

superimpose on one another due to the influence of component tolerances and measurement errors. Reference [5] considers the optimization method of the frequency of the excitation signal. This assures that the parameter loci intersect themselves at possibly large angles and they have similar lengths. Unfortunately, these improvements did not eliminate the disadvantages of the method, particularly of superimposing parameter loci. The work in reference [2] presented 3D (dimension) and 4D methods of fault localization and identification in the linear circuit shown in Fig.1. And Zbigniew and Romuald in reference [3] proposed the idea of transferring the parameter loci family by the 3-D (dimension), 4-D and 6D methods.

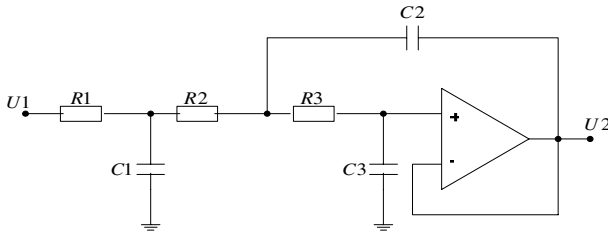


Fig. 1. Third-order low-pass Butterworth filter, where $R_1 = R_2 = R_3 = 5k\Omega$, $C_1 = 44.5nF$, $C_2 = 110nF$, $C_3 = 6.42nF$

Successful though they are, the method mentioned above can not be applied to fault diagnosis of nonlinear circuits, while the nonlinear nature of the problem widely exists [6-8], for example, reference [8] refers to that a parameter value changes by a certain factor, but the responses do not change by the same factor, which leads to the relationship between the circuit responses and the component characteristics is nonlinear, even though the circuit may be linear. And, in general, the nonlinear characteristics of linear and nonlinear circuits result from the situations when not all the nodes of the circuit under test (CUT) are accessible. The work in reference [7] proposed a closed form representation for section-wise piecewise-linear models. Based on this, we apply Katzenelson's piecewise-linear analysis [6] to nonlinear networks in this work.

The methods based on space transformation can be easily implemented in neural networks working as fault feature selection techniques. This makes application of the methods developed by Zbigniew and Romuald [2-3] to this area very appealing. By far the most popular neural networks architecture is the back-propagation neural network (BPNN) [9]. So, in this paper, we use BPNNs for fault location. However, a BP neural network successfully trained for given samples is not guaranteed to provide desired associations for untrained inputs as well [12]. Concerning this problem some authors showed experimentally that the generalization capability could remarkably be enhanced by training the network with noise injected inputs [12-13]. The work in reference [12] mathematically explains why and how the noise injection to inputs has such effect. In the context of this work, we aim to present a systematic way to diagnose faulty behavior in nonlinear circuits. First, space transformation listed in [2-3], will be briefly reviewed later, and used to select candidate features from circuits under test. Next, an approach using genetic algorithm-based programming concept is

developed to deal with component tolerances which influence the generalization of BPNNs. Finally, a neural network is trained and tested for fault diagnosis of nonlinear circuits.

The material in this paper is arranged in the following order. In section 2 we briefly review the bilinear transformation in multi-space proposed in reference [2-3]. Section 3 discusses the PWL models of nonlinear components and describes the fault feature extraction techniques in this paper. Section 4 covers the principals of obtaining the training patterns of neural networks for fault diagnosis with tolerances. And a detailed discussion of the BPNN used in analog fault diagnosis is provided in this Section too. Section 5 covers the diagnostic examples and results. And the conclusions are given in section 6.

2 Bilinear Transformation in Multi-Space

In this paper, multi-space transformations refer to the transformations of 3D and $2m(m \geq 2)$ spaces. For single fault diagnosis with 3D methods, the voltage transfer functions K_u and input admittance Y_{in} are measured. They can be expressed in the form of bilinear functions of p_i -parameters:

$$H_i^1(p_i) = \frac{A_i^1 p_i + B_i^1}{C_i^1 p_i + D_i^1} \tag{1}$$

$$H_i^2(p_i) = \frac{A_i^2 p_i + B_i^2}{C_i^2 p_i + D_i^2} \tag{2}$$

Two transformations using both functions above are listed below:

$$T_i(p_i) = \text{Re}(H_i^1(p_i))\mathbf{i} + \text{Im}(H_i^1(p_i))\mathbf{j} + |H_i^2(p_i)|\mathbf{k} \tag{3}$$

$$V_i(p_i) = \text{Re}(F_i^1(p_i))\mathbf{i} + \text{Im}(F_i^1(p_i))\mathbf{j} + \text{Re}(F_i^2(p_i))\mathbf{k} + \text{Im}(F_i^2(p_i))\mathbf{l} \tag{4}$$

Where $\mathbf{i}, \mathbf{j}, \mathbf{k}, \mathbf{l}$ are versors (unit vectors compliant with axes), $|\cdot|$ is absolute value, $\text{Re}(\cdot)$ and $\text{Im}(\cdot)$ are real and imaginary parts of the circuit function, respectively. The transformations (3) and (4) map changes of p_i -loci in 3D and 4D space, respectively. Also, this transformation can be extended to $2m$ space, which has the following form:

$$T_i^s(p_i) = \sum_{j=1}^s (\text{Re}(F_i^j(p_i))\mathbf{k}^{2j-1} + \text{Im}(F_i^j(p_i))\mathbf{k}^{2j}) \tag{5}$$

Where $F_i^j(p_i)$ is j th function about the parameter p_i ; $j = 1, 2, \dots, s$, s is the number of the functions. When $s = 3$, transformation (5) corresponds to the transformation in 6D space.

Experiments are carried out when the values of components for C_2 and R_2 in Fig.1 change in the range $0.1p_{inorm} \rightarrow 10p_{inorm}$, where p_{inorm} is the nominal value of the

component. The distances between C_2 and R_2 -loci for 2D, 3D, 4D and 6D methods are compared in Fig.2. It is seen that for the 6D method based on transformation (5), the p_i -loci distances are much greater with better implications for its efficiency. The increase of the distance between C_2 and R_2 -loci is 24.97 times greater than for the 2D method, 2.087 times greater than for the 3D method and 1.523 times greater than for the 4D method. It is shown that distances between C_2 and R_2 -loci are separated, and distances between them are considerably greater than on a plane. This is an advantage of the multi-spaces based method, which leads to better fault resolution as well as to robustness against the influence of component tolerances and measurement errors.

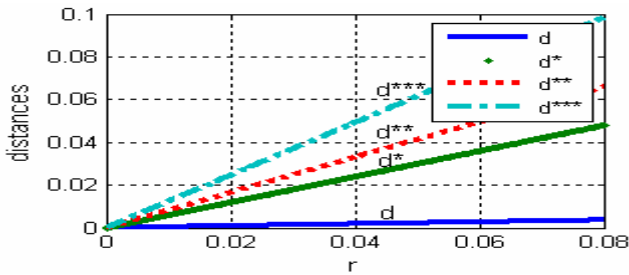


Fig. 2. Distances between C_2 and R_2 -loci taken at the same normalized radius (r) from a nominal point, as a function of radius-value for a 2D (d), 3D (d^*), 4D (d^{**}), and 6D (d^{***})

The method mentioned above in references [2-3] can only be applied to fault diagnosis of linear circuits, while the nonlinear nature of the problem widely exists [6-10]. Thus, there is a need for extension of this method in nonlinear circuits. In this paper, we represent nonlinear circuit elements by their piecewise linear models.

3 Feature Selections for Nonlinear Analog Fault Diagnosis

3.1 The PWL Models of Nonlinear Components

The idea of PWL is presented in Fig.3. Assume that nonlinear components are restricted to those that are voltage-controlled, that is, with a characteristic $i = G(v)$. Its PWL characteristic is composed of several linear segments in the $i-v$ plane. The c -th operating region is defined by the admittance, say Y_c . So we can replace the nonlinear element with its PWL admittance Y_c . A Y_c -segment PWL model can be used to approximate the nonlinear characteristic of the nonlinear element Y_c . Once all nonlinear elements are replaced in such way, the nodal equation formulated can be solved with Katzenelson's algorithm proposed in reference [6].

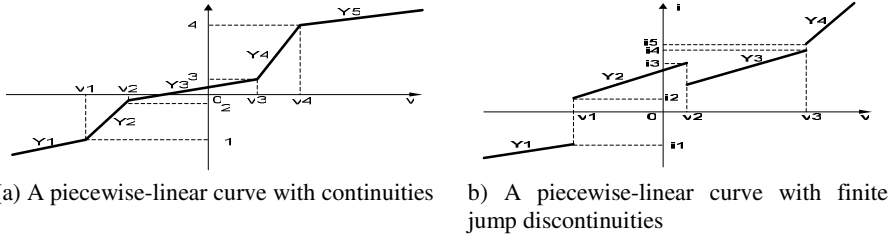


Fig. 3. Piece-wise linear characteristics

If the nonlinear characteristic of the nonlinear element is not varied, that is, the nonlinear element is fault free. But the operating region of the nonlinear element can be shift from Y_j to Y_k ($j, k \in Y_c, c \in Z$) (here Z is the total number of the operating regions of the nonlinear element) because of the other faulty elements in CUT. Thus, after the nonlinear components being replaced by their PWL model, the space transformation techniques mentioned in reference [2-3] can be extended to nonlinear circuits.

3.2 The Fault Feature Extraction Techniques

The space transformation techniques mentioned above can be easily implemented in neural networks working as fault feature selection techniques, which make the application of the methods to fault diagnosis very appealing. In applying space transformation to fault feature selection, one needs to perform the transformation on (1) or (2) to preserve the real and imaginary parts and magnitudes of the circuit functions. This means that the real and imaginary parts and/or magnitudes of the circuit functions constitute the inputs of neural networks without considering tolerances.

Concerning the generalization capability of neural networks and the influence of component tolerances on the circuits' response, we train the neural networks with noise injected inputs and optimize the response of the circuits by their tolerance limits.

Let h_i^r be the range of r -th circuit function in i -th node due to component tolerances. Then it can be expressed as

$$\begin{aligned} & \max \quad \left| h_i^r \right| \\ & \text{s.t.} \quad x_{j \min} \leq x_j \leq x_{j \max}; j = 1, 2, \dots, N \end{aligned} \tag{6}$$

Where x_j is the parameter of the j -th element in the circuit with its tolerance range of $[x_{j \min}, x_{j \max}]$ and N is the total number of elements in the circuit. The optimization can be realized conveniently based on the conventional genetic algorithm [13]. Thus the train and validate pattern groups are

$$XX + \left| h_i^r \right| \cdot \text{rand}(\cdot) \tag{7}$$

Where $\text{rand}(\cdot)$ is random noise. And the feature vectors XX are extracted according to space transformation through (5).

4 Neural Network-Based Analog Fault Diagnosis of Nonlinear Circuits

4.1 BPNNs Overview

The BP neural network is a kind of multilayer, feed forward network [1] [13]. The architecture of three-layer BPNN is shown in Fig.4.

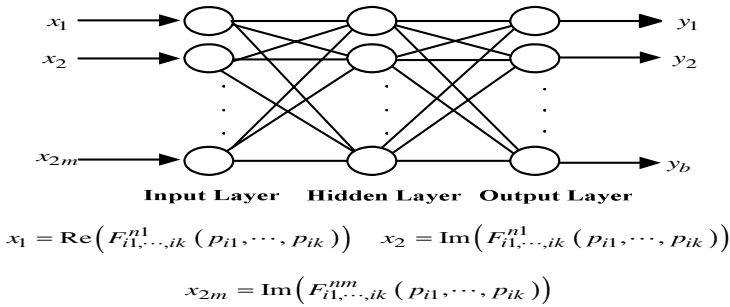


Fig. 4. Three-layer BPNN for 3D or 2m-D method

In Fig. 4, this corresponds to setting the number of output layer neurons b equals to the number of fault classes. And the number of input layer neurons equals to $2m$. The output for this architecture would be given by

$$y_l = f_2 \left(\sum_{j=1}^{S2} \left(\omega_j \cdot f_1 \left(\sum_{i=1}^{2m} (\omega_{ji} \cdot x_i + b_j) \right) + b_l \right) \right) \tag{8}$$

In this equation, $S2$ is the number of hidden-layer neurons. ω_{ji} represents the weight between j -th neuron in hidden-layer and i -th inputs and ω_j is the weight between l -th neuron in output-layer and j -th neuron in hidden-layer. b_j and b_l are the biases of hidden-layer and output-layer neurons respectively and $l = 1, 2, \dots, b$ is the number of output-layer neurons.

The activation function of the hidden layer which is a sigmoid has the following form:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{9}$$

The activation function of the output layer could be sigmoid or linear, here we choose linear function. The reason for this is that the output will be limited to a certain range with using tanh function, but using the linear function the output without such limitation. Hence the activation function of the output is linear having the form as $f(X) = X$. The error function for this architecture would be the sum of squares error given by

$$E = \frac{1}{2} \sum_{l=1}^b (t_l - y_l)^2 \quad (8)$$

In this equation, t_l and y_l represent the target and actual outputs of the neural network respectively.

4.2 Neural Network-Based Analog Fault Diagnosis of Nonlinear Circuits

BPNNs have the advantages of large-scale parallel processing, parallel storing, robust adaptive learning, and on-line computation [1]. These advantages make the application of neural networks to analog fault diagnosis very appealing. In this approach, all the faults are modeled by a unique set of features that the network learns during the training phase. These features, together with the associated fault classes, are presented to the network as input-output pairs. The network is then allowed to adjust its weight and bias parameters to learn the desired input-output relationship. Next, the network is presented with a set of features as input during the testing phase and determines the fault class. According to the foregoing discussion the algorithm for identification of the faulty elements can be enumerated:

- (1) Replace the nonlinear elements with their PWL models and obtain the equivalent linear circuit of the CUT.
- (2) Perform sensitivity analysis of the circuit under test (CUT) to select the most suitable test nodes and a set F of faulty elements considered as possible faulty is formed.
- (3) For an arbitrary single faulty elements belonging to F, decide the circuit functions, for example, the voltage transfer functions K_u and/or input admittance Y_{in} , etc., then the fault features are selected by equation (5). The train and validate patterns are available based on equation (7) and the NN are trained to its equilibrium point. Prove the trained NN with validate patterns.
- (4) At the test stage, we measure the circuit functions, then diagnosis pattern are extracted according to equation (5) And the test patterns are available based on equation (7).
- (5) Pass the normalized test pattern to the trained NN and the output of NN declares the faults of CUT.
- (6) Evaluate the diagnosis results.

5 Simulation Results

In this section, in order to verify our proposed method, we use the Matlab7 and PSpice to simulate a nonlinear circuit shown in Fig.5 and illustrate the method for fault feature selection by space transformations and the analog fault diagnosis technique developed above. As in Fig.5, the volt-ampere property of the nonlinear voltage-controlled resistor is $i(\nu) = 0.1\nu + 0.02\nu^2$. The tolerances of these elements are 10%.

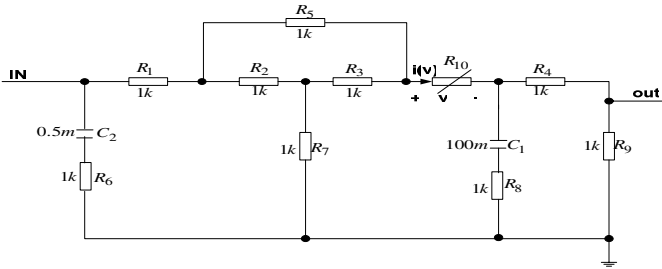


Fig. 5. A nonlinear circuit

The sinusoid stimulus with 235Hz is fed into the node *IN* , and 4D and 6D based neural networks techniques for fault diagnosis are taken here. The PWL model of nonlinear resistor was created by applying Chebychev’s approximation theory based on alternation theorem [8], that is, the function $i(v)$ can be approximated by $p(v) = a_n v_n + b_n$ with a maximum predefined error $|(i(v)-p(v))/i(v)| \leq \alpha$. Where α is a prespecified number, here $\alpha = 0.05$.Thus a 7-segment PWL model was used to approximate $i(v) = 0.1 v + 0.02 v^2$ when $v \in [-3.5v, 3.5v]$.The 7 segments was list in Table1.

Table 1. The 7 segments PWL model of the nonlinear resistor

No.	$g^{(l)}$ (mS)	$I^{(l)}$ (mA)	Corner Points(v_n & v_{n+1})
1	0.5286	0.7245	-3.500& -1.7895
2	0.2109	0.0950	-1.7895& -0.892
3	1.0000	0.0000	-0.8920 &-0.3050
4	0.1012	0.0000	-0.3050 & 0.3050
5	0.1342	-0.0145	0.3050&1.1565
6	0.2698	-0.1816	1.1565& 2.1791
7	0.5889	-0.9053	2.1791&3.500

After the nonlinear component is substituted by its 7 segments PWL models, the space transformation technique can be used in these segments.

For comparison, 2D, 3D, 4D and 6D methods are implemented in neural networks. We study the following faults:

The actual parameters for components $R_1 \sim R_9$ and $C_1 \sim C_2$ are $0.5 p_{inorm}$, $0.7 p_{inorm}$, $0.9 p_{inorm}$, $1.1 p_{inorm}$, $1.3 p_{inorm}$, $1.5 p_{inorm}$, $2 p_{inorm}$, $3 p_{inorm}$, $5 p_{inorm}$, $7 p_{inorm}$ and $10 p_{inorm}$, where p_{inorm} is the nominal value.

The property of the nonlinear resistor is transferred to $i(v) = 5v + 0.08v^2$.

(1) Methods based on 2D transformation: Measure the real and the imaginary part of the voltage transfer function $H^1(p_i) = T_u = u_{out}/u_{in}$ or input admittance function

$H^2(p_i) = Y_{in}$. Then the train and validate pattern groups are obtained on equation (7) after the tolerance is calculated using the conventional genetic algorithm [13]. And the number of training pattern is 61, while the number of test pattern is 51. We assume that the goal of prespecified error is 0.01. Using the structures of 2-32-20-12 for $2D/T_u$ method and 2-41-31-12 for $2D/Y_{in}$ method to diagnose the faults mentioned above, we obtain that the average accuracies is 49% and 58%, respectively. Here the number of output neurons of the neural networks includes one for no-fault class. The output of NN declares the faults: the actual faults are the ones that the fault types declare if the output of NN is approximated enough to one of the fault type in the predefined dictionary; otherwise, the fault element is the nonlinear resistor. It is shown that the 2D method does not give satisfied results for fault localization.

(2) Methods based on 4D and 6D transformations: For both the 4D and 6D methods the same circuit functions are used: the voltage transfer function $H^1(p_i) = T_u = u_{out}/u_{in}$, and the input admittance $H^2(p_i) = Y_{in}$. Additionally the current-voltage transmittance $H^3(p_i) = T_{iu} = i_{out}/u_{in}$ for the 6D method. Then the train and validate pattern groups are obtained on equation (7) after the tolerance was calculated using the conventional genetic algorithm [13]. And the number of training pattern is 61, while the number of test pattern is 51. We assume that the goal of prespecified error is 0.01. Using the structures of 4-23-18-12 for 4D method and 6-15-14-12 for 6D method to diagnose the faults mentioned above, we obtain that the average accuracies is 80% and 95%, respectively. The advantages from the fact that in the 4D and 6D neural networks we obtain increasing separation of fault classes. They give a higher probability of correct fault localization.

6 Conclusions

An efficient neural network-based fault diagnosis approach to nonlinear analog circuits using circuit function transformations in multiple spaces and normalization as preprocessors is proposed. After the nonlinear components are replaced by their PWL models, the space transformation approach proposed in references [2-3] can be applied to analysis of nonlinear analog circuits. And the advantages of the methods, following from greater distances between component parameters locus in space, are better fault localization resolution and robustness against the influence of component tolerances and measurement errors. Our study indicates that these techniques lead to a satisfactory performance in nonlinear analog fault diagnosis.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant No.50677014, Doctoral Special Fund of Ministry of Education under grant No. 20060532002, High-Tech Research and Development Program of China (No.2006AA04A104), the Program for New Century Excellent Talents in University of China (NCET-04-0767), Foundation of Hunan Provincial Science and Technology (06JJ2024).

References

1. Aminian, M., Collins, H.W.: Analog Fault Diagnosis of Actual Circuits Using Neural Networks. *IEEE Trans. on Instrum. and Meas.* 51, 544–550 (2002)
2. Czaja, Z., Zielonko, R.: Fault Diagnosis in Electronic Circuits Based on Bilinear Transformation In 3-D and 4-D Spaces. *IEEE Trans. On Instrum. and Meas.* 52, 97–102 (2003)
3. Czaja, Z., Zielonko, R.: On Fault Diagnosis of Analogue Electronic Circuits Based on Transformations in Multi-Dimensional Spaces. *Measurements* 35, 293–301 (2004)
4. Martens, G., Dyck, J.: Fault Identification in Electronic Circuit with the Aid of Bilinear Transformation. *IEEE Trans. Reliability* 2, 99–104 (1972)
5. Czaja, Z.: The Fault Location Algorithm Based on Two Circuit Functions. In: *Proc. XVI Imeko World Congress, Vienna, Austria, vol. VI*, pp. 29–32 (2000)
6. Katzenelson, J.: An Algorithm for Solving Nonlinear Resistor Networks. *Bell Syst. Tech. J.* 44, 553–556 (1965)
7. Chua, L.O., Sung, M.: Section-Wise Piecewise-Linear Functions: Canonical Representation, Properties, and Applications. *Proceeding of the IEEE* 65, 915–929 (1977)
8. Bandler, J.W., Salama, A.E.: Fault Diagnosis of Analog Circuits. *Proceedings of the IEEE* 73, 1279–1325 (1985)
9. Deng, Y., He, Y.: On the Application of Artificial Neural Networks to Fault Diagnosis in Analog Circuits with Tolerances. In: *Proceedings of ICSP 2000*, pp. 1639–1642 (2000)
10. He, Y., Sun, Y.: A Neural-Based Nonlinear L1-Norm Optimization Approach for Fault Diagnosis of Nonlinear Circuits with Tolerance. *IEE proceedings circuits, Devices and systems* 148, 223–228 (2001)
11. He, Y., Tan, Y., Sun, Y.: Wavelet Neural Network Approach for Fault Diagnosis of Analog Circuits. *IEE Proc. -Circuits Devices Systems* 151, 379–384 (2004)
12. Matsuoka, K.: Noise Injection into Inputs in Back-Propagation Learning. *IEEE Trans. On Systems, Man, and Cybernetics* 22, 436–440 (1992)
13. McInerney, M., Dhawan, A.P.: Use of Genetic Algorithms with Back Propagation in Training of Feed-Forward Neural Networks. *Proc. IEEE ICNN*, 203–208 (1993)

An Intelligent Fault Diagnosis Method Based on Multiscale Entropy and SVMs

Long Zhang¹, Guoliang Xiong², Hesheng Liu³, Huijun Zou¹,
and Weizhong Guo¹

¹ School of Mechanical Eng., Shanghai Jiaotong University, Shanghai 200240, China

² School of Mechatronic Eng., East China Jiaotong University, Nanchang 330013, China

³ Department of physics, Shangrao Normal College, Shangrao 334001, China

longzh@126.com

Abstract. Sample entropy (SampEn) has been applied in many literatures as a statistical feature to describe the regularity of a time series. However, as components of mechanical system usually interact and couple with each other, SampEn may cause inaccurate or incomplete description of a mechanical vibration signal due to the fact that SampEn is calculated at only one single scale. In this paper, a new method, named multiscale entropy (MSE), taking into account multiple time scales, was introduced for feature extraction from fault vibration signal. MSE in tandem with support vector machines (SVMs) constitutes the proposed intelligent fault diagnosis method. Details on the parameter selection of SVMs were discussed. In addition, performances between SVMs and artificial neural networks (ANNs) were compared. Experiment results verified the proposed model.

Keywords: Fault diagnosis, Sample entropy, Multiscale entropy, SVMs.

1 Introduction

On line machine condition monitoring and fault diagnosis has been increasingly attracting attention from the research and engineering community worldwide over the past decades [1]. Generally, a simple condition monitoring system is approached from a pattern classification perspective. It can be decomposed into three general steps: data acquisition, feature extraction, and condition classification, among which the latter two are of significant importance.

Due to instantaneous variations in friction, damping, stiffness or loading conditions, mechanical systems are often characterized by non-linear behaviors that in turn make the vibration signals complex and non-linear. As such, commonly used signal processing techniques including time and frequency domain techniques, as well as advanced signal processing techniques, such as wavelet transform and time-frequency representation, may all exhibit limitations. Therefore, techniques for non-linear dynamic parameter estimation provide a good alternative to extracting defect-related features hidden in the complex and non-linear vibration signals [1, 2]. Hitherto, a number of non-linear parameter identification techniques have been investigated and introduced to fault diagnosis, among which correlation dimension is a typical one [3, 4, and 5].

Reliable estimation of correlation dimension, however, usually requires very long data set that is difficult or even impossible to be achieved especially in on-line, real-time monitoring and diagnosis. A brief review on non-linear dynamic parameters used for feature extraction and fault diagnosis can be found in literature [1], and in the same literature appropriate entropy (ApEn) was introduced and selected as a tool for rolling bearing health monitoring. Although ApEn has found its ways in fields of physiological signal and machine vibration signal processing [1,2,6], however due to the bias within its estimation, ApEn is heavily dependent on the data length and its estimated value is uniformly lower than that expected for short records, and lacks relative consistency as well [7]. In order to overcome the shortcomings of ApEn, Richman and Moorman [7] proposed a new kind of entropy, named sample entropy (SampEn) which seems much more promising and has attracted a lot of attention [7,8].

In a relatively recent paper [9], a new entropy based measure of complexity, which is the multiple scale entropy (MSE), was introduced. The authors applied their new complexity measure (i.e. MSE) to distinguish between young healthy hearts and congestive heart failure. Moreover, the MSE was able to distinguish atrial fibrillation from healthy hearts [9, 10, and 11]. The key to the MSE method lies in a multiscale approach [12]. Consider a machine composed of gears, bearings, shafts and other mechanical components [13]. Even a modest amount of machine complexity will result in measured vibration signals that contain multiple intrinsic oscillatory modes due to the interaction of the mechanical components, that implies non-linear dynamic parameters applied on single scale (such as ApEn and SampEn of original time series) may be insufficient for characterizing machine vibration signals. For this reason, the multiscale method was introduced and tried in the present study in the hope of improving performances of traditional non-linear dynamic methods on single scale within the context of machine fault diagnosis. To the best of the authors' knowledge, the MSE has not been applied in the field of fault diagnosis so far. Its advanced properties attract us for a trial of its use.

After extracting MSE acting as feature vectors, one needs a classifier to fulfill automated fault recognition. A number of intelligent classification algorithms, such as artificial neural networks (ANNs) and support vector machines (SVMs) have been successfully applied to automated machine fault diagnosis [14]. The main advantages of SVMs lie in the fact that it can perform better in the processing of small-sample sized learning problems and has better generalization due to the replacement of Empirical Risk Minimization used in ANNs by Structural Risk Minimization. Due to these merits, SVMs has become a new research hotspot in recent years and have been applied successfully in many domains. Hence, SVMs was selected in the present study as a fault classifier.

2 Theoretical Background

2.1 Multiscale Entropy (MSE)

MSE was computed according to the procedure published by Costa et al [9, 10, and 11]. Given a one-dimensional discrete time series, $\{x_1, \dots, x_i, \dots, x_N\}$, one can constructed

consecutive coarse-grained time series $\{y^{(\tau)}\}$ determined by the scale factor τ , according to the equation

$$y_j^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i$$

where τ represents the scale factor and $1 \leq j \leq N/\tau$. In other words, coarse-grained time series for scale τ are obtained by taking arithmetic mean of τ neighboring original values without overlapping (Fig.1). The length of each coarse-gained time series is N/τ . For scale 1, the coarse-grained time series is simply the original time series. Then SampEn or ApEn is computed for the coarse-gained time series at each scale and plotted as a function of the scale factor.

SampEn is a refinement of traditionally used regularity measure ApEn statistics. Details on the SampEn algorithm can be found in many literatures [15]. Briefly, SampEn quantifies the regularity of time series. It reflects the condition probability that two sequences of m consecutive data points, which are similar to each other (within give tolerance r), will remain similar when one consecutive point is included [15]. The SampEn algorithm underlying the MSE computation requires setting two parameters: the tolerance level r and the pattern length m . According to previous studies, it has been chosen that $r = 0.15 \times$ standard deviation of the time series to avoid distortion of SampEn values by changes in signal magnitude and $m = 2$.

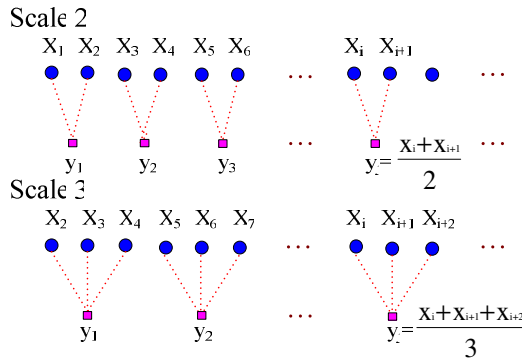


Fig. 1. The scheme illustrating the coarse-graining of an original time series for scales $\tau = 2$ and $\tau = 3$

2.2 Support Vector Machines (SVMs)

SVM is a classification method derived from Statistical Learning Theory (SLT) by Vapnik and Chervonenkis [16]. Its basic idea is to map the original data to a higher dimensional feature space and find the optimal hyperplane in the space that maximizes the margin between the classes, as illustrated in Fig.2. The essential difference between SVMs and ANNs lies in their requirements imposed on the hyperplane. In the case of SVMs, it is desired to find the hyperplane with maximal margin and minimal class error ratio on training data, whereas in ANNs, only the latter is necessary. According to the

SLT, for a trained classifier to predict unseen samples, the actual risk consists of two parts, i.e. empirical risk (R_{emp}) and confidence interval ϕ .

$$R \leq R_{emp} + \phi$$

As such, in order to minimize actual risk, the only satisfaction with minimal class error of training data, i.e. minimal R_{emp} , is not enough. In ANNs, large number of training data is required to minimize ϕ . Training data, however, is usually limited, especially for fault samples of machinery. Therefore, the requirement on maximal margin is taken into account in SVMs to account for ϕ . According to the SLT, maximal margin will result in minimal ϕ . Hence, SVMs doesn't rely much on the amount of training data and possesses advantages over ANNs with respect to generalization performance. Maximal margin in conjunction with minimal error of training data is referred to as minimal structural risk.

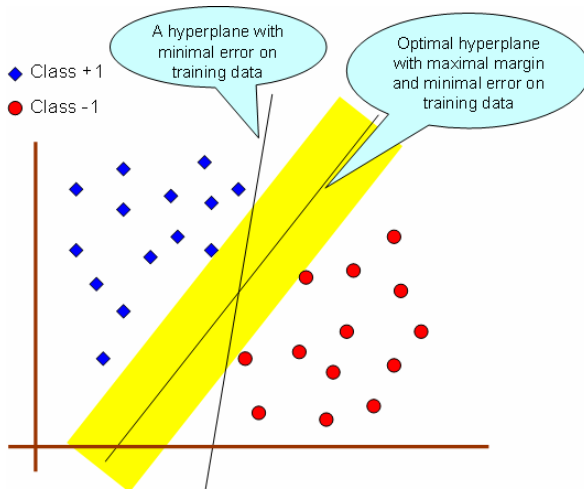


Fig. 2. The line with a yellow background illustrating an optimal hyperplane

3 Experimental Analysis

3.1 Experimental Setup

In order to validate the proposed fault diagnosis method, experimental analyses on rolling element bearings were conducted. All the bearing data and related system analyzed in this paper belong to Case Western Reserve Lab [17].

The test stand, shown in Fig.3, consists of a 2 hp, three-phase induction motor (left), a torque sensor (middle) and a dynamometer (right) connected by a self-aligning coupling (middle). The dynamometer is controlled so that desired torque load levels can be achieved. The test bearings support the motor shaft at both the drive end and fan end. Single point faults were introduced to the test bearings using electro-discharge

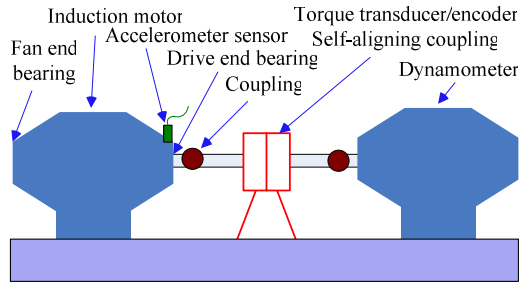


Fig. 3. Bearing fault test stand

machining with fault diameters of 7 mil, (1 mil=0.001 inches). Vibration data was collected using accelerometers, which were attached to the housing with magnetic bases [17, 18].

Vibration signals of drive end bearing under 0 hp load collected from good, outer race fault, inner race fault and ball fault condition were analyzed. For each condition, there are 25 samples and each sample contains 4096 data points. The sampling frequency is 12,000Hz, and the approximate motor speed is 1797 rpm. Hence, motor rotates about 11 revolutions over the time interval of 4096 data points. One sample of the four conditions was shown in Fig.4.

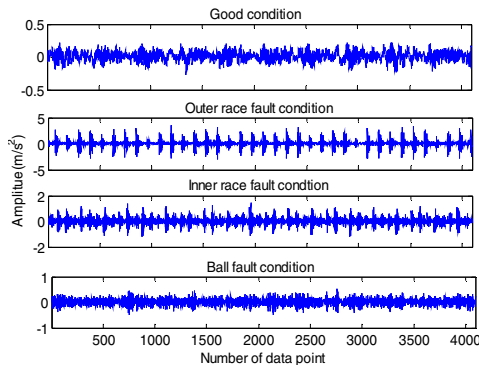


Fig. 4. Vibration signals in time domain of four different bearing conditions

3.2 Calculation of MSE

MSE, in essence, is to calculate sample entropy (or other type of entropy like ApEn) over a set of scales. For this purpose, prior to the calculation of MSE, there are three parameters to be defined, i.e. the tolerance level r , the pattern length m and the maximal scale factor. Values of r and m have been determined in section 2.1 according to previous studies. Maximal scale factor was selected as 50 by experiments. Figure 5 shows the MSE of the samples depicted in figure 4, from which the four conditions can't be separated linearly, so a nonlinear classifier is necessary.

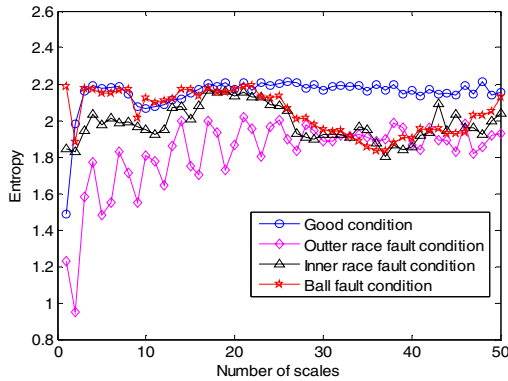


Fig. 5. MSE calculated over 50 time scales for the signals shown in Fig.4

3.3 Classification with SVMs

SVMs classify data in the form like a linear function. When linearly inseparable data are concerned, SVMs make use of Kernel trick to map the original data to a higher dimensional space where the data may be linearly separable. There are various kernel functions used in SVMs, such as linear, polynomial, radial basis function (RBF) and sigmoid kernel. Since RBF kernel has less hyperparameters and less numerical difficulties, it is a reasonable first choice [19].

Basic SVMs is developed for binary classification. In practice, however, there are many scenarios involving multi-class classification. To this end, a lot of methods have been developed such as one-against-rest and one-against-one. For the case of one-against-rest, there are possibly some data that can't be classified into any classes or will be classified into many classes. To avoid this deficiency, one-against-one paradigm was adopted.

After the determination of the type of kernel function and the multi-class method, there are still two parameters to be determined, i.e. penalty parameter C and RBF width parameter γ . This can be solved by cross-validation and grid-search [19]. As stated above, there are 25 samples for each bearing condition respectively. Among them are randomly selected 10 samples as training data, remainder 15 samples as testing data. Because of less training data, a two-fold cross-validation was implemented to determine the C and γ . For a given value of C and γ , the 10 training samples were split into two subsets each containing 5 samples. Then, the second subset was predicted by the SVMs trained with the first subset, and vice versa. The sum of the two prediction accuracy was used as a performance metric to evaluate the given C and γ . The values of C and γ within a prescribed range ($C \in 2^{-5:15}$, $\gamma \in 2^{-15:13}$ in this paper) achieving the highest prediction accuracy will be selected for future applications [19]. The validation prediction accuracies (testing accuracy) of the total $21 \times 19 = 399$ pairs of C and γ are shown in Fig. 6; where there are 373 cases reaching a testing accuracy of 100%. So many optimal values make it confused for the selection. In practice, if there are no other more training data available, any pair of the optimal C and γ is a possible candidate. In order to examine the performances of

all the possible 373 candidates, their classification rates on testing data are shown in Fig. 7. Among all the cases, there are 35.92% getting a rate of 100%, and 45.83% getting a rate of 98.33%, 11.26% getting a rate of 96.67, as well as 4.29% getting the lowest rate of 93.33%. As such, all of the 373 candidates produced very promising results in the classification of testing data. The details on the classification regarding the testing data of three couples of C and γ representing three kinds of classification accuracy are depicted in Table 1. For all the three cases, there are no samples of fault bearing misclassified into good condition, which implies a small risk of the proposed fault diagnosis method.

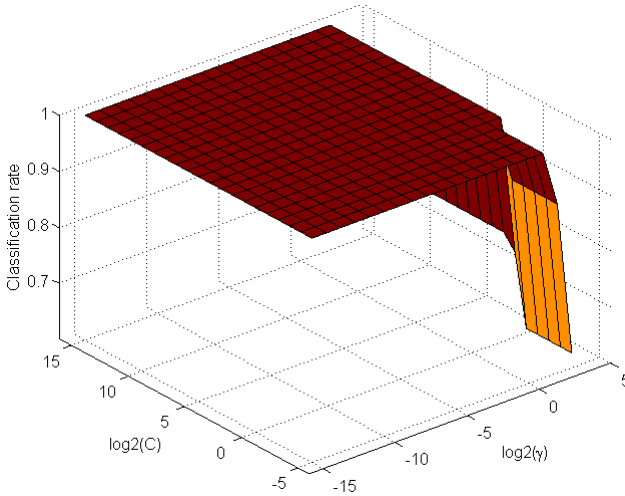


Fig. 6. Classification rate of various pairs of C and γ in cross-validation

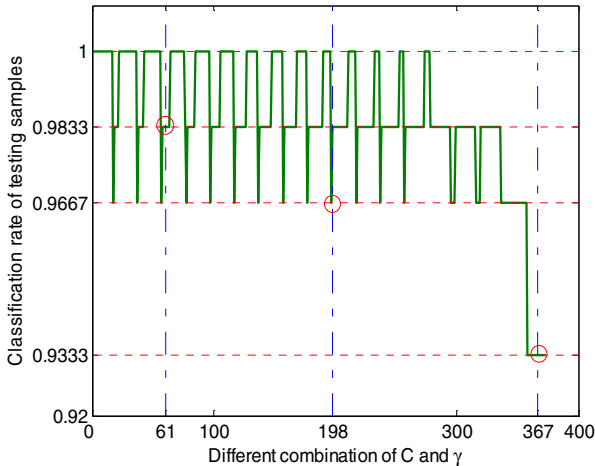


Fig. 7. Classification rate on testing data with various pairs of C and γ from the optimal values determined by cross-validation

A three-layer BP neural network with 80 and 4 nodes in middle and output layer respectively was also investigated. As showed in Table 2, a total of 11 samples were misclassified with a classification rate of 81.67%, among which 6 fault samples were treated as good condition. This will lead to a rather larger risk than SVMs. Proper increase of the node number of middle layer will give a possible raise to the accuracy. Nevertheless, due to the high dimension of feature characteristics (i.e. MSE over 50 scales), too many nodes in middle layer will render the training and testing speed very slow, which doesn't suit on-line and real-time application.

Table 1. Confusing Matrix of SVMs with different values of C and γ indicated by lines vertical to the horizontal axis in Fig.7 at points 61, 198 and 367

	$C = 2^{13}, \gamma = 2^{-13}$				$C = 2^3, \gamma = 2^{-6}$				$C = 2^9, \gamma = 2^2$			
Condition	A	B	C	D	A	B	C	D	A	B	C	D
A	14	0	1	0	13	0	2	0	12	1	2	0
B	0	15	0	0	0	15	0	0	0	15	0	0
C	0	0	15	0	0	0	15	0	0	0	15	0
D	0	0	0	15	0	0	0	15	0	0	1	14
Accuracy	98.33%				96.67%				93.33%			

A--Good condition, B--Outer race fault, C--Inner race fault, D--Ball fault.

Table 2. Confusing Matrix of a three-layer BP network

Condition	A	B	C	D
A	15	0	0	0
B	0	15	0	0
C	2	2	11	0
D	4	0	3	8
Accuracy	81.67%			

A--Good condition, B--Outer race fault, C--Inner race fault, D--Ball fault.

4 Conclusions

Experiments verified the effectiveness of the combination of multiscale entropy (MSE) and SVM. MSE can extract the nonlinear information hidden in vibration signals over multiple scales. SVMs are superior in terms of good generalization performance as well as less dependence on the amount of training data. In comparison with SVMs, the accuracy rate of BP network is slightly lower. The rather higher accuracy of both SVMs and BP in turn demonstrated the effectiveness of the features extracted by MSE. How to determine the maximal scale factor to which MSE will be calculated is an open question. In this work, it's selected as 50 by trials.

Acknowledgments. The work was supported by the Natural Science Foundation of Jiangxi Province under the grant No. 0450017.

References

1. Yan, R.Q., Gao, R.X.: Approximate Entropy as a Diagnostic Tool for Machine Health Monitoring. *Mechanical Systems and Signal Processing* 21, 824–839 (2007)
2. Yan, R.Q., Gao, R.X.: Machine Health Diagnosis Based on Approximate Entropy. In: *ICMT 2004 Instrumentation and Measurement Technology Conference*, pp. 2054–2059. IEEE Press, Italy (2004)
3. Jiang, J.D., Chen, J.: The Application of Correlation Dimension in Gearbox Condition Monitoring. *Journal of Sound and Vibration* 224, 529–541 (2004)
4. Logan, D., Mathew, J.: Using the Correlation Dimension for Vibration Fault Diagnosis of Rolling Element Bearings-I. Basic Concepts. *Mechanical Systems and Signal Processing* 10(3), 241–250 (1996)
5. Logan, D., Mathew, J.: Using the Correlation Dimension for Vibration Fault Diagnosis of Rolling Element Bearings-II. Selection of Experimental Parameters. *Mechanical Systems and Signal Processing* 10(3), 251–260 (1996)
6. Xu, Y.G., Li, L.J., He, Z.J.: Approximate Entropy and Its Applications in Mechanical Fault Diagnosis. *Information and Control* 31(6), 547–551 (2002)
7. Richman, J.S., Moorman, J.R.: Physiological Time-Series Analysis Using Approximate Entropy and Sample Entropy. *Am. J. Physiol. H.* 278, 2039–2049 (2000)
8. Haitham, M.A., Alan, V.S.: Use of Sample Entropy Approach to Study Heart Rate Variability in Obstructive Sleep Apnea Syndrome. *IEEE Trans. Bio. Eng.* 50, 1900–1904 (2007)
9. Costa, M., Goldberger, A.L., Peng, C.K.: Multiscale Entropy Analysis of Complex Physiologic Time Series. *Phys. Res. Lett.* 89(6), 68–102 (2002)
10. Costa, M., Goldberger, A.L., Peng, C.K.: Multiscale Entropy to Distinguish Physiologic and Synthetic RR Time Series. *Computers in Cardiology* 29, 137–140 (2002)
11. Costa, M., Goldberger, A.L., Peng, C.K.: Multiscale Entropy Analysis of Biological Signals. *Phys. Rev. E.* 71, 21906 (2005)
12. Ranjit, A.T., Georg, A.G.: On Multiscale Entropy Analysis for Physiological Data. *Physica A* 366, 323–332 (2006)
13. Fan, X.F., Zuo, M.J.: Machine Fault Feature Extraction Based on Intrinsic Mode Functions. *Meas. Sci. Technol.* 19, 245105(12pp) (2008)
14. Lu, S., Yu, F.J., Liu, J.: Bearing Fault Diagnosis Based on K-L Transform and Support Vector Machine. In: *3rd IEEE International Conference on Natural Computation*, pp. 522–527. IEEE Press, New York (2007)
15. Trunkvalterova, Z., Javorka, M., Tonhajzerova, I., Javorkova, J., Lazarova, Z., Javorka, K., Baumert, M.: Reduced Short-Term Complexity of Heart Rate and Blood Pressure Dynamics in Patients with Diabetes Mellitus Type 1: Multiscale Entropy Analysis. *Physiol. Meas.* 29, 817–828 (2008)
16. Vapnik, V.N.: *The Nature of Statistical Learning Theory*. Springer, Berlin (1995)
17. Case Western Reserve University Bearing Data Center Website, http://www.eecs.case.edu/laboratory/bearing/welcome_overview.htm
18. Lou, X.S., Loparo, K.A.: Bearing Fault Diagnosis Based on Wavelet Transform and Fuzzy Inference. *Mechanical Systems and Signal Processing* 18, 1077–1095 (2004)
19. Hsu, C.W., Chang, C.C., Lin, C.J.: *A Practical Guide to Support Vector Classification* (2008), <http://www.csie.ntu.edu.tw/~cjlin>

Multi-objective Robust Fault Detection Filter Design in a Finite Frequency Range

Yu Cui, Xin-han Huang, and Min Wang

Department of Control Science and Engineering,
Huazhong University of Science and Technology, Wuhan 430074, China
cuiyu211@hotmail.com

Abstract. The robust fault detection filter (RFDF) design problem for linear time invariant (LTI) system with unknown inputs is studied. The design objectives are set to minimize a combined performance index H_∞ / H_- over the specified frequency range as well as to satisfy regional constraints on filter poles. A linear matrix inequality (LMI) based solution is proposed with the aid of the recently developed generalized KYP lemma. The advantage of the proposed solution lies in that the real values of the H_∞, H_- indices over the given finite frequency range are accessible during the solving process. Based on that, an optimal solution minimizing the performance index H_∞ / H_- with filter poles lie in the specified region is proposed by converting the design problem to a standard LMI optimization problem. An aircraft design example demonstrates the effectiveness of the proposed solution.

Keywords: Fault detection, Filter design, LMI.

1 Introduction

With an increasing demand for higher performance as well as for more safety and reliability of dynamic systems, fault detection (FD) has received more and more attention. The most popular FD techniques are model-based FD schemes which are usually realized in two stages: residual generation and decision making. The residual generation estimates the output of the system through observations and compares the real output and the estimated as the residual. The decision making analyzes the residual and makes the decision concerning the presence and location of faults. However, the unknown inputs such as model errors, uncertain disturbances, and process and measurement noises may result in significant changes in the residual, leading to false alarm. So it does not only require the residual remain sensitive to the fault signal, but also keep robust against the unknown inputs.

To solve this problem, H_∞ / H_- paradigm is introduced to the RFDF design. The H_∞ norm is used to enforce robustness of the residual against unknown inputs while H_- index is introduced to measure the fault sensitivity of the residual. H_∞ / H_- based RFDF design method is first proposed with H_- index defined at $\omega=0$ [1] or over non-zero frequency range [2,7]. Then the definition of H_- index is extended in [3,4] to represent the true worst-case fault sensitivity of the residual to the fault. By adopting

this new definition, iterative algorithms are proposed for the RFDF design in [5,6]. However, the values of H_∞, H_- indices over the finite frequency range obtained in these papers are all approximate values. Meanwhile, the proposed design methods can not guarantee an optimal solution for making the combined performance index H_∞ / H_- smallest.

In practice, it is often the case that the energy of the fault lies in a certain finite frequency range. To deal this situation, a new RFDF design method for LTI system with unknown inputs is proposed in this paper. The design objectives include the minimization of the combined performance index H_∞ / H_- over a specified finite frequency range, which is set to detect the fault of smallest energy possible, as well as regional constraints on filter poles for the consideration of the transient performance of the filter. Compared with the aforementioned work, the real value of the H_∞, H_- indices of the filter over the specified finite frequency range can be obtained with the aid of the newly developed KYP lemma, which is helpful for improving the performance of the designed filter. Based on that, an optimal solution to the filter design problem is given by converting it to a standard LMI based optimal problem.

Notation: For a matrix M , Its transpose, complex conjugate transpose and the Moore-Penrose inverse are denoted by M^T, M^*, M^\dagger . The Hermitian part of a square matrix M is denoted by $He(M) := M + M^*$. For matrices Q and P , $Q \otimes P$ means their Kronecker product.

2 Problem Formulation

Consider a LTI system described by

$$\Sigma: \dot{x} = Ax + Bu + B_f f + B_d d, \tag{1}$$

$$y = Cx + C_f f + C_d d, \tag{2}$$

where $x \in \mathbb{R}^n$ is the state; $u \in \mathbb{R}^{n_u}$ is the control input; $y \in \mathbb{R}^{n_y}$ is the measured output. $f \in \mathbb{R}^{n_f}$ is the fault signal that belong to the frequency range $\Omega := [\omega_1 \ \omega_2]$ ($\omega_2 > \omega_1 \geq 0$) which can be system component faults, actuator faults or sensor faults. $d \in \mathbb{R}^{n_d}$ is the unknown input including modeling errors, unknown disturbance, process and measurement noises. $A, B, C, B_f, B_d, C_f, C_d$ are known constant matrix with appropriate dimensions. Without loss of generality, assume (C, A) is observable and d is in the same frequency range as f since the components of the unknown input d over other frequency range can be removed by setting a post band-through filter after the residual output.

The fault detection filter used to generate residual is of the form:

$$\begin{aligned} \dot{\hat{x}} &= A\hat{x} + Bu + L(y - \hat{y}) \\ y &= C\hat{x} \\ r &= y - \hat{y} \end{aligned}, \tag{3}$$

where L is the of filter gain matrix to be designed., r the residual output. Define the error state as $e = x - \hat{x}$, it follows from (1),(2),(3) that the error and residual dynamic system can be described as

$$\begin{aligned} \dot{e} &= (A - LC)e + (B_f - LC_f)f + (B_d - LC_d)d \\ r &= Ce + C_f f + C_d d \end{aligned} \quad (4)$$

Let $\bar{A} = A - LC$, $\bar{B}_f = B_f - LC_f$, $\bar{B}_d = B_d - LC_d$, then we have

$$G_{rd} = C(sI - \bar{A})^{-1} \bar{B}_d + C_d, \quad G_{rf} = C(sI - \bar{A})^{-1} \bar{B}_f + C_f,$$

where G_{rd} is the transfer function matrix from d to r , G_{rf} is the transfer function matrix from f to r . Then the RFDF design problem can be formulated as follows:

RFDF design problem. The goal of the RFDF filter design problem is to find a suitable filter gain matrix, such that the residual generated by system (3) satisfies the following requirements:

(R.1) The poles of the filter lie in a specified region of the open left-half complex plane.

(R.2) $\|G_{rf}(j\omega)\|_{-}^{\Omega} > \gamma$, where $\|G_{rf}\|_{-}^{\Omega}$ is the minimum singular value of the transfer function matrix G_{rf} over the finite frequency range Ω .

(R.3) $\|G_{rd}(j\omega)\|_{\infty}^{\Omega} < \beta$, where $\|G_{rd}\|_{\infty}^{\Omega}$ is the maximum singular value of the transfer function matrix G_{rd} over the finite frequency range Ω .

(R.4) the robustness/sensitivity ratio (noise-signal ratio) β / γ is minimized

Remark 1. Requirement (R.1) is introduced to tune the transient response of the residual. This feature is very important, since in the latter decision making step the residual is usually post-processed by a hypothesis based test to make a final decision of the fault. Requirement (R.2) is used to guarantee the worst-case sensitivity of the residual to fault signal in the H_{-} sense. Since the residual r is not only depends on the fault signal f but also on the unknown input d , the effect of the unknown input of the system on the residual can greatly increase under some circumstances, thus considerably degrade the fault detection performance. The requirement (R.3) is introduced here to represent the worst-case robustness of the residual to the unknown input. It is clear that the smaller γ / β in (R.4) is, the better the fault detection performance of the fault detection filter will be. In fact, if the fault detection objective is to achieve low false alarm rate and the threshold is set, then the value

$$J_s = 2 \inf(\|G_{rd}(s)\|_{\infty}^{\Omega} / \|G_{rf}\|_{-}^{\Omega})$$

gives the smallest fault signal that is guaranteed to be detected [7].

3 LMI Solutions of the RFDF Design Problem

In this section, a solution of the RFDF design problem based on the LMI techniques is presented. The fault sensitivity, unknown input robustness specifications as well as filter poles regional constraint are all converted to the corresponding LMI formulations. Based on that, the RFDF design problem is transformed into a standard LMI optimization problem aimed at minimizing the combined performance index H_{∞} / H_{-} .

3.1 LMI Solution of the (R.1) Specification

For the convenience of the fault decision, it is necessary to assign the poles of the fault detection filter in a suitable region of the open left-half complex plane. A significant region $S(a, r', \theta)$ (see Fig.1) is the clipped sector described by

$$S(a, r', \theta) := \{x + jy \in C \mid x < -a, |x + jy| < r', \text{tg } \theta < -\frac{|y|}{x}\}.$$

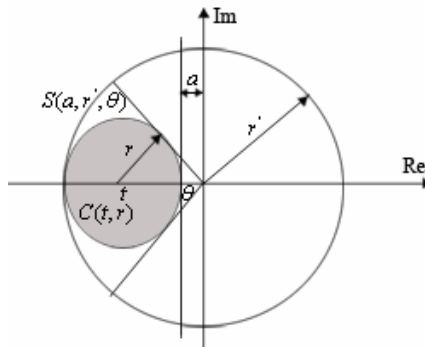


Fig. 1. Sector $S(a, r', \theta)$ and disk $C(t, r)$

The control system with poles assigned in this region can achieve a satisfactory transient response. For instance, let $\omega_{1,2} = -\zeta\omega_n \pm j\omega_d$ be the complex poles of a second-order system. If $\omega_{1,2} \in S(a, r', \theta)$, then it follows that the system acquires the minimum decay rate of α , the minimum damping ratio of $\zeta = \cos \theta$ and the maximum damped natural frequency of $\omega_d = r \sin \theta$, which further enforces different bounds on the other transient response performance indices, such as rising time, settling time and overshoot.

However, note that the region $S(a, r', \theta)$ has three characteristic equations, so it becomes complicated when formulated in terms of LMIs. Besides, in that case the calculation load of the RFDF design turns out to be huge. To solve this problem, the disk

$$C(t, r) := \{x + jy \in C \mid |x + jy - t| < r\},$$

(see Fig.1) with its boundary inscribing $S(a, r', \theta)$ is used instead.

Before giving the LMI solution of poles assignment in disk $C(t, r)$, lemma 1 is given below.

Lemma 1[8]. For any matrix A , let Φ be any given Hermitian matrix satisfying $\det(\Phi) < 0$, $\sigma(G, T) := \begin{bmatrix} G \\ I \end{bmatrix}^* T \begin{bmatrix} G \\ I \end{bmatrix}$, then the following statements are equivalent:

- i) Each eigenvalue λ of A satisfies $\sigma(\lambda, \Phi^T) < 0$
- ii) There exist matrix X and $P = P^* > 0$ such that

$$\Phi \otimes P < \text{He} \begin{bmatrix} -I \\ A \end{bmatrix} X \begin{bmatrix} -qI & pI \end{bmatrix}, \tag{5}$$

where $r = [p \ q]^*$ is any arbitrary fixed vector satisfying $r^* \Phi r < 0$

It is clear from lemma 1 that the eigenvalues of A can be confined in the specified region defined by appropriate Φ . In particular, if Φ is set as

$$\Phi := \begin{bmatrix} 1 & -t \\ -t & |t|^2 - r^2 \end{bmatrix}, (t < 0) \tag{6}$$

it defines the disk $C(t, r)$. Substituting (6) into (5), we can get following Corollary.

Corollary 1: Consider system (4), the eigenvalues of \bar{A} lie in the disk $C(t, r)$ if and only if there exist matrix X and $P_0 = P_0^T > 0$ such that

$$\begin{bmatrix} P_0 & -tP_0 \\ -tP_0 & (t^2 - r^2)P_0 \end{bmatrix} < \text{He} \begin{bmatrix} -X \\ A^T X - C^T Y \end{bmatrix} \begin{bmatrix} -qI & pI \end{bmatrix}, \tag{7}$$

where $Y = L^T X$ and p, q is any fixed real numbers satisfying

$$p^2 - 2tpq + q^2(t^2 - r^2) < 0.$$

3.2 LMI Solution of the (R.2) and (R.3) Specification

Since the fault signal f lies in the finite frequency range Ω , it is beneficial for the RFDF design if the real value of the H_∞, H_- indices over Ω can be obtained. Fortunately, this goal can be achieved with the aid of the generalized KYP lemma [9] which gives the exact LMI characterizations of the H_∞, H_- indices in a finite frequency interval for continuous setting. Based on that, the LMI solutions of the requirement (R.2) and (R.3) can be derived.

Let $\omega_c = \frac{1}{2}(\omega_1 + \omega_2)$, then it is easy to get the following corollaries from the generalized KYP lemma.

Corollary 2. Consider system (4), let $\Pi_1 = \begin{bmatrix} -I & 0 \\ 0 & \gamma^2 I \end{bmatrix}$, then $\|G_{fj}(j\omega)\|_-^\Omega > \gamma$ if and only if there exist symmetrical matrices P_1 and $Q_1 > 0$ satisfying

$$\begin{bmatrix} \bar{A} & \bar{B}_f \\ \mathbf{I} & 0 \end{bmatrix}^* \Phi \begin{bmatrix} \bar{A} & \bar{B}_f \\ \mathbf{I} & 0 \end{bmatrix} + \begin{bmatrix} \mathbf{C} & \mathbf{C}_f \\ 0 & \mathbf{I} \end{bmatrix}^* \Pi_1 \begin{bmatrix} \mathbf{C} & \mathbf{C}_f \\ 0 & \mathbf{I} \end{bmatrix} < 0, \quad (8)$$

where $\Phi_1 = \begin{bmatrix} -\mathbf{Q}_1 & \mathbf{P}_1 + j\omega_c \mathbf{Q}_1 \\ \mathbf{P}_1 - j\omega_c \mathbf{Q}_1 & -\omega_1 \omega_2 \mathbf{Q}_1 \end{bmatrix}$.

Corollary 3. Consider system (4), let $\Pi_2 = \begin{bmatrix} \mathbf{I} & 0 \\ 0 & \beta^2 \mathbf{I} \end{bmatrix}$, then $\|G_{rd}(j\omega)\|_\infty^\Omega < \beta$ if and only if there exist symmetrical matrices \mathbf{P}_2 and $\mathbf{Q}_2 > 0$ satisfying

$$\begin{bmatrix} \bar{A} & \bar{B}_d \\ \mathbf{I} & 0 \end{bmatrix}^* \Phi \begin{bmatrix} \bar{A} & \bar{B}_d \\ \mathbf{I} & 0 \end{bmatrix} + \begin{bmatrix} \mathbf{C} & \mathbf{C}_d \\ 0 & \mathbf{I} \end{bmatrix}^* \Pi_2 \begin{bmatrix} \mathbf{C} & \mathbf{C}_d \\ 0 & \mathbf{I} \end{bmatrix} < 0, \quad (9)$$

where $\Phi_2 = \begin{bmatrix} -\mathbf{Q}_2 & \mathbf{P}_2 + j\omega_c \mathbf{Q}_2 \\ \mathbf{P}_2 - j\omega_c \mathbf{Q}_2 & -\omega_1 \omega_2 \mathbf{Q}_2 \end{bmatrix}$.

Note that (8), (9) is not convex due to the product terms between \mathbf{L}, \mathbf{P}_i and \mathbf{Q}_i , necessary transformations are needed. Consider Corollary 2 first, with the Finsler’s Lemma [10], we have following lemma which is essential for the LMI solution of (R.2). Since the derivation of the lemma 2 is similar to the lemma [3] in [11], it is omitted here in the interest of space.

Lemma 2. Consider system (4), suppose symmetrical matrices $\mathbf{P}_1, \mathbf{Q}_1 > 0$ and $\mathbf{R}_1 \in \mathbf{R}^{n \times (2n+n_f+n_y)}$, then the following statements are equivalent:

i) There exist a gain matrix \mathbf{L} such that (R.1) and

$$(\mathbf{S}\mathbf{R}_1^T)^\perp \mathbf{S}\mathbf{T} \begin{bmatrix} \Phi & 0 \\ 0 & \Pi_1 \end{bmatrix} \mathbf{T}^T \mathbf{S}^T (\mathbf{S}\mathbf{R}_1^T)^\perp < 0$$

hold where $\psi = [\mathbf{I} \ 0], \mu_1 = \begin{bmatrix} \bar{A}^T & \mathbf{C}^T \\ \bar{B}_f^T & \mathbf{C}_f^T \end{bmatrix}, \mathbf{S} = \begin{bmatrix} \mu_1 & \mathbf{I} \\ \psi & 0 \end{bmatrix}$.

ii) There exist $\chi \in X(\psi, \mathbf{R}_1), \mathbf{Y} = \mathbf{L}^T \mathbf{X}$ such that

$$\mathbf{T} \begin{bmatrix} \Phi & 0 \\ 0 & \Pi_1 \end{bmatrix} \mathbf{T}^T < \text{He} \begin{bmatrix} -\chi & \\ \Psi \chi + \Lambda \mathbf{Y} \mathbf{R}_1 & \end{bmatrix},$$

where $\Lambda = \begin{bmatrix} -\mathbf{C}^T \\ -\mathbf{C}_f^T \end{bmatrix}, \Psi = \begin{bmatrix} \mathbf{A}^T & \mathbf{C}^T \\ \mathbf{B}_f^T & \mathbf{C}_f^T \end{bmatrix}$.

\mathbf{T} is a permutation matrix such that

$$[\mathbf{M}_1 \ \mathbf{M}_2 \ \mathbf{M}_3 \ \mathbf{M}_4] \mathbf{T} = [\mathbf{M}_1 \ \mathbf{M}_3 \ \mathbf{M}_2 \ \mathbf{M}_4],$$

$$X(\psi, \mathbf{R}_1) := \{\psi^\dagger \mathbf{X} \mathbf{R}_1 + (\mathbf{I} - \psi^\dagger \psi) \mathbf{V} \mid \mathbf{X} \in \mathbf{R}^{n \times n}, \det \mathbf{X} \neq 0, \mathbf{V} \in \mathbf{R}^{(n+n_y) \times (2n+n_y+n_f)}\} .$$

Substituting $R_1 = [I \ 0 \ I \ -B_f]$ into Lemma 2, the following theorem gives the LMI solution of the requirement (R.2).

Theorem 1. Consider system (4), if there exist symmetrical matrices P_1 and $Q_1 > 0$ satisfying

$$\begin{bmatrix} -B_f^T Q_1 B_f - C_f^T C_f + \gamma^2 I & -B_f^T Q_1 - B_f^T \Theta + C_f^T C \\ \bullet & -2P_1 - \lambda Q_1 - C^T C \end{bmatrix} < 0, \tag{10}$$

where $\Theta = P_1 + j\omega_c Q_1$, $\lambda = 1 - \omega_1 \omega_2$, then the following statements are equivalent:

- i) $\|G_{rf}(s)\|_{\infty}^{\Omega} > \gamma$,
- ii) There exist matrices X and $Y = L^T X$ such that

$$\begin{bmatrix} -Q_1 + X + X^T & V_1^T & \Theta + X - X^T A - V_1^T C + Y^T C & -XB_f - X^T B_f - V_1^T C_f + Y^T C_f \\ \bullet & -I + V_2^T + V_2 & V_3 - V_2^T C & \Theta & V_4 - V_2^T C_f \\ \bullet & \bullet & -\alpha\omega\omega_c Q_1 - \Delta_1 - \Delta_1^T & A^T X B_f - C^T V_4 - C^T Y B_f - X^T B_f - V_3^T C_f + Y^T C_f \\ \bullet & \bullet & \bullet & \gamma^2 I - \Delta_2 - \Delta_2^T \end{bmatrix} < 0, \tag{11}$$

where $V = [V_1 \ V_2 \ V_3 \ V_4]$, $V_1 \in \mathbb{R}^{n_s \times n}$, $V_2 \in \mathbb{R}^{n_s \times n_s}$, $V_3 \in \mathbb{R}^{n_s \times n}$, $V_4 \in \mathbb{R}^{n_s \times n_{f_s}}$, $\Delta_1 = A^T X + C^T V_3 - C^T Y$, $\Delta_2 = -B_f^T X B_f + C_f^T V_4 + C_f^T Y B_f$.

The LMI solution of the requirement (R.3) can be established easily following the similar way to that of the requirement (R.2), so it is just given immediately.

Theorem 2. Consider system (4), if

$$C_d^T C_d - \beta^2 I < 0, \tag{12}$$

then $\|G_{rd}(s)\|_{\infty}^{\Omega} < \beta$ holds if and only if there exist symmetrical matrices P_2 and $Q_2 > 0$, X , $Y = L^T X$ such that

$$\begin{bmatrix} -Q_2 & V_a^T & P_2 + j\omega_c Q_2 + X - V_a^T C & -V_a^T C_d \\ \bullet & I + V_b^T + V_b & V_c - V_b^T C & V_d - V_b^T C_d \\ \bullet & \bullet & -\alpha\omega\omega_c Q_2 - \Delta_3 - \Delta_3^T & -C^T V_d - X^T B_d - V_c^T C_d + Y^T C_d \\ \bullet & \bullet & \bullet & -\beta^2 I - V_d^T C_d - C_d^T V_d \end{bmatrix} < 0, \tag{13}$$

where $\Delta_3 = A^T X - C^T Y + C^T V_c$, $\tilde{V} = [V_a \ V_b \ V_c \ V_d]$, $V_a \in \mathbb{R}^{n_s \times n}$, $V_b \in \mathbb{R}^{n_s \times n_s}$, $V_c \in \mathbb{R}^{n_s \times n}$, $V_d \in \mathbb{R}^{n_s \times n_d}$

3.3 LMI Solution of RFDF Design Problem

Now the fault sensitivity, unknown input robustness specifications as well as pole assignment requirements are all converted to the corresponding LMI formulations. Based on that, the noise-signal ratio minimization problem (R.4) can be solved by the following theorem.

Theorem 3. Consider system (4), RFDF design problem has the following solution:

$$\max \bar{r}$$

s.t.

$$\begin{bmatrix} -\mathbf{B}_f^T \tilde{\mathbf{Q}}_1 \mathbf{B}_f - \hat{\mathbf{r}} \mathbf{C}_f^T \mathbf{C}_f + \bar{r} \mathbf{I} & -\mathbf{B}_f^T \tilde{\mathbf{Q}}_1 - \mathbf{B}_f^T \Theta + \hat{\mathbf{r}} \mathbf{C}_f^T \mathbf{C} \\ \bullet & -2\tilde{\mathbf{P}}_1 - \lambda \tilde{\mathbf{Q}}_1 - \hat{\mathbf{r}} \mathbf{C}^T \mathbf{C} \end{bmatrix} < 0$$

$$\begin{bmatrix} -\tilde{\mathbf{Q}}_1 + \tilde{\mathbf{X}} + \tilde{\mathbf{X}}^T & \tilde{\mathbf{V}}_1^T & \Theta + \tilde{\mathbf{X}} - \tilde{\mathbf{X}}^T \mathbf{A} - \tilde{\mathbf{V}}_1^T \mathbf{C} + \tilde{\mathbf{Y}}^T \mathbf{C} & -\tilde{\mathbf{X}} \mathbf{B}_f - \tilde{\mathbf{X}}^T \mathbf{B}_f - \tilde{\mathbf{V}}_1^T \mathbf{C}_f + \tilde{\mathbf{Y}}^T \mathbf{C}_f \\ \bullet & -\hat{\mathbf{r}} \mathbf{I} + \tilde{\mathbf{V}}_2^T + \tilde{\mathbf{V}}_2 & \tilde{\mathbf{V}}_3 - \tilde{\mathbf{V}}_2^T \mathbf{C} & \tilde{\mathbf{V}}_4 - \tilde{\mathbf{V}}_2^T \mathbf{C}_f \\ \bullet & \bullet & -\alpha \omega \tilde{\mathbf{Q}}_1 - \tilde{\Delta}_1 - \tilde{\Delta}_1^T & \mathbf{A}^T \tilde{\mathbf{X}} \mathbf{B}_f - \mathbf{C}^T \tilde{\mathbf{V}}_4 - \mathbf{C}^T \tilde{\mathbf{Y}} \mathbf{B}_f - \tilde{\mathbf{X}}^T \mathbf{B}_f - \tilde{\mathbf{V}}_3^T \mathbf{C}_f + \tilde{\mathbf{Y}}^T \mathbf{C}_f \\ \bullet & \bullet & \bullet & \bar{\mathbf{r}} \mathbf{I} - \tilde{\Delta}_2 - \tilde{\Delta}_2^T \end{bmatrix} < 0$$

$$\hat{\mathbf{r}} \mathbf{C}_d^T \mathbf{C}_d - \mathbf{I} < 0$$

$$\begin{bmatrix} -\tilde{\mathbf{Q}}_2 & \tilde{\mathbf{V}}_a^T & \tilde{\mathbf{P}}_2 + j\omega \tilde{\mathbf{Q}}_2 + \tilde{\mathbf{X}} - \tilde{\mathbf{V}}_a^T \mathbf{C} & -\tilde{\mathbf{V}}_a^T \mathbf{C}_d \\ \bullet & \hat{\mathbf{r}} \mathbf{I} + \tilde{\mathbf{V}}_b^T + \tilde{\mathbf{V}}_b & \tilde{\mathbf{V}}_c - \tilde{\mathbf{V}}_b^T \mathbf{C} & \tilde{\mathbf{V}}_d - \tilde{\mathbf{V}}_b^T \mathbf{C}_d \\ \bullet & \bullet & -\alpha \omega \tilde{\mathbf{Q}}_2 - \tilde{\Delta}_3 - \tilde{\Delta}_3^T & -\mathbf{C}^T \tilde{\mathbf{V}}_d - \tilde{\mathbf{X}}^T \mathbf{B}_d - \tilde{\mathbf{V}}_c^T \mathbf{C}_d + \tilde{\mathbf{Y}}^T \mathbf{C}_d \\ \bullet & \bullet & \bullet & -\mathbf{I} - \tilde{\mathbf{V}}_d^T \mathbf{C}_d - \mathbf{C}_d^T \tilde{\mathbf{V}}_d \end{bmatrix} < 0$$

$$\begin{bmatrix} \tilde{\mathbf{P}}_0 & -t\tilde{\mathbf{P}}_0 \\ -t\tilde{\mathbf{P}}_0 & (t^2 - r^2)\tilde{\mathbf{P}}_0 \end{bmatrix} < \text{He} \begin{bmatrix} -\tilde{\mathbf{X}} \\ \mathbf{A}^T \tilde{\mathbf{X}} - \mathbf{C}^T \tilde{\mathbf{Y}} \end{bmatrix} \begin{bmatrix} -q\mathbf{I} & p\mathbf{I} \end{bmatrix},$$

where $\hat{r} = \beta^{-2}$, $\bar{r} = (\gamma / \beta)^2$, symmetrical matrix $\tilde{\mathbf{P}}_i > 0 \ i = 0, 1, 2$, $\tilde{\mathbf{Q}}_i > 0 \ i = 1, 2$. $\Theta = \tilde{\mathbf{P}}_1 + j\omega_c \tilde{\mathbf{Q}}_1$, $\Delta_1 = \mathbf{A}^T \tilde{\mathbf{X}} + \mathbf{C}^T \tilde{\mathbf{V}}_3 - \mathbf{C}^T \tilde{\mathbf{Y}}$, $\Delta_2 = -\mathbf{B}_f^T \tilde{\mathbf{X}} \mathbf{B}_f + \mathbf{C}_f^T \tilde{\mathbf{V}}_4 + \mathbf{C}_f^T \tilde{\mathbf{Y}} \mathbf{B}_f$, $\Delta_3 = \mathbf{A}^T \tilde{\mathbf{X}} - \mathbf{C}^T \tilde{\mathbf{Y}} + \mathbf{C}^T \tilde{\mathbf{V}}_c$ and the filter gain matrix is given by $\mathbf{L} = (\tilde{\mathbf{Y}} \tilde{\mathbf{X}}^{-1})^T$.

Proof. Pre- and post multiply (7),(10),(11),(12),(13) by diagonal matrix Γ with appropriate dimensions whose diagonal elements are set as β^{-1} . Let $\tilde{\mathbf{X}} = \hat{\mathbf{r}} \mathbf{X}$, $\tilde{\mathbf{Y}} = \hat{\mathbf{r}} \mathbf{Y}$, $\tilde{\mathbf{P}}_i = \hat{\mathbf{r}} \mathbf{P}_i \ i = 0, 1, 2$, $\tilde{\mathbf{Q}}_i = \hat{\mathbf{r}} \mathbf{Q}_i \ i = 1, 2$, $\tilde{\mathbf{V}}_i = \hat{\mathbf{r}} \mathbf{V}_i \ i = 1, 2, 3, 4, a, b, c, d$. Substitute these relations into (7), (10), (11), (12), (13) then we can get the expression above.

Remark 2. Note that theorem 3 is formulated as a standard optimization problem, so the filter gain matrix can be easily obtained by solving this problem using MATLAB. It is obvious that the design objective β / γ is minimized if $(\gamma / \beta)^2$ is maximized.

4 Numerical Example

In this section, a numerical example is presented to illustrate the effectiveness of the proposed LMI based multi-objective RFDF design method. Consider the linearized longitudinal dynamics model of an aircraft given in [12], its state-space equations are given as:

$$\dot{x} = Ax + Bu + B_f f + B_d d, \quad y = Cx + C_d d,$$

with

$$A = \begin{bmatrix} -0.08 & -0.03 & -0.157 & 0 \\ -0.73 & -0.377 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & -8.65 & 0 & -0.5 \end{bmatrix}, \quad B = B_f = \begin{bmatrix} 1.54 & -0.020 \\ -0.10 & -0.056 \\ 0 & 0 \\ 0 & -6.50 \end{bmatrix}, \quad B_d = \begin{bmatrix} 0 \\ 0.3 \\ 0.2 \\ 0.05 \end{bmatrix},$$

$$C = I_{4 \times 4}, \quad C_d = [0.2 \quad 0.1 \quad 0 \quad 0.3]^T,$$

where, $x = [\nu \quad \alpha \quad \theta \quad q]$, ν is the mach number, α is the attack angle, θ is the pitch angle, $q = \dot{\theta}$ is the pitch rate. $u = [\zeta_p \quad \zeta_e]$, ζ_p is the throttle input, ζ_e is the elevon deflection. Assume the fault signal f is in the frequency range $[0 \quad 0.1]$.

To improve transient behavior of the residual, we require the poles of the fault detection filter are confined in the disk $C(-4, 3.5)$. According to Corollary 1, set $p = 1$, $q = -1$. Applying theorem 3, we get the fault detection filter gain matrix:

$$L = \begin{bmatrix} 0.81348 & -1.6703 & 0.48232 & 0.01446 \\ -0.68074 & 1.3815 & 0.15624 & 0.99331 \\ -0.05885 & -0.90627 & 2.5238 & 1.008 \\ 0.45019 & -1.717 & -2.0327 & 0.43889 \end{bmatrix},$$

with $\gamma = 1.5356$, $\beta = 0.3742$. It can be observed that the poles of the filter

$$\lambda_{1,2} = -1.000, \lambda_3 = -1.9086, \lambda_4 = -2.2061$$

are all confined in the specific region $C(-4, 3.5)$.

The behavior of the residual during the simulation is depicted in Fig.2. The unknown input d is simulated as the white noise with noise power 0.001. The actuator fault is simulated as $f = [1 \quad 1]^T$, $t \geq 5s$. It is clear that the designed fault detection filter exhibits a good unknown input decoupling property and maintains a satisfactory fault sensitivity performance. It can be observed that the fault will be detected shortly after its occurrence. Due to the suitable pole assignment, the residual curve is smooth and stable, thus greatly improving the validity of the fault decision.

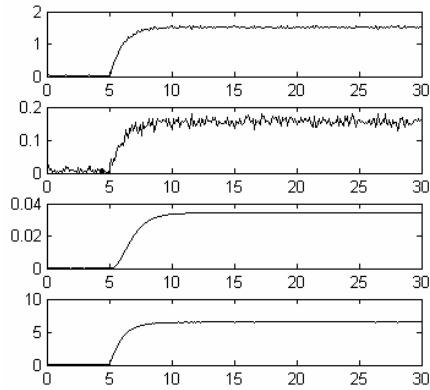


Fig. 2. Behavior of the residuals in simulation

5 Conclusion

In this paper, a general RFDF design method for LTI systems with unknown inputs is proposed taking into account multiple objectives. The objectives include robustness/sensitivity ratio specification and regional constraint on filter poles. Since the fault signals usually have energy in a finite frequency range, the robustness/sensitivity ratio is formulated in terms of H_∞/H_- index over the finite frequency. The multi-objective RFDF design problem is converted into a standard convex optimization problem in terms of LMIs. Through solving this multi-objective optimization problem, the optimal RFDF gain matrix can be determined. A numerical example has demonstrated the effectiveness of the proposed method.

References

1. Hou, M., Patton, R.J.: An LMI Approach to H_∞/H_- Fault Detection Observers. In: UKACC International Conference on Control, UK, pp. 305–310 (1996)
2. Ding, S.X., Jeinsh, T., Frank, P.M., Ding, E.L.: A Unified Approach to the Optimization of Fault Detection Systems. *International Journal of Adaptive Control and Signal Processing* 14(7), 725–745 (2000)
3. Liu, J., Wang, J.L., Yang, G.H.: Worst-Case Fault Detection Observer Design: An LMI Approach. In: *The International Conference on Control and Automation*, pp. 1243–1247 (2002)
4. Liu, J., Wang, J.L., Yang, G.H.: An LMI Approach to Minimum Sensitivity Analysis with Application to Fault Detection. *Automatica* 41(11), 1995–2004 (2005)
5. Wang, H.B., Lam, J., Ding, S.X., Zhong, M.Y.: Iterative Linear Matrix Inequality Algorithms for Fault Detection with Unknown Inputs. *Journal of Systems and Control Engineering* 219(2), 161–172 (2005)
6. Wang, H.B., Wang, J.L., Lam, J.: Robust Fault Detection Observer Design. *Iterative LMI Approaches* 129(1), 77–83 (2007)
7. Rank, M., Niemann, H.: Norm Based Design of Fault Detectors. *International Journal of Control* 72(9), 773–783 (1999)

8. Iwasaki, T., Hara, S.: Robust Control Synthesis with General Frequency Domain Specifications: Static Gain Feedback Case. In: American Control Conference, pp. 4613–4618 (2004)
9. Iwasaki, T., Hara, S.: Generalized KYP Lemma: Unified Frequency Domain Inequalities with Design Applications. IEEE Transaction of Automatic Control 50(1), 41–59 (2005)
10. Skelton, R.E., Iwasaki, T., Grigoriadis, K.M.: A Unified Algebraic Approach to Linear Control Design. Taylor&Francis, London (1997)
11. Wang, H., Wang, J.L.: Fault Detection Observer Design in Low Frequency Domain. In: IEEE International Conference on Control Applications, Singapore, pp. 976–981 (2007)
12. Wang, J.H., Shi, Z.K., Cao, L.: Application of Mixed H_2/H_∞ Robust Control to Aircraft. Flight Dynamics 18(4), 61–64 (2000)

Intelligent Technique and Its Application in Fault Diagnosis of Locomotive Bearing Based on Granular Computing

Zhang Zhousuo, Yan Xiaoxu, and Cheng Wei

Department of Mechanical Engineering, State Key Laboratory
for Manufacturing Systems Engineering
Xi'an Jiaotong University, Xi'an 710049, P.R. China

zzs@mail.xjtu.edu.cn, daokyman@hotmail.com, xjtuwei@gmail.com

Abstract. This paper presents a new approach to intelligent fault diagnosis of the machinery based on granular computing. The tolerance granularity space mode is constructed by means of the inner-class distance defined in the attributes space. Different features of the vibration signals, including time domain statistical features and frequency domain statistical features, are extracted and selected using distance evaluation technique as the attributes to construct the granular structure. Finally, the proposed approach is applied to fault diagnosis of locomotive bearings, testing results show that the proposed approach can reliably recognize different faulty categories and severities.

Keywords: Granular computing, Tolerance relations, Granularity structure, Fault diagnosis.

1 Introduction

Nowadays, as the large-scale complex system, the high-speed locomotive is playing a very important role in the development of the society. And the normal operation of the rolling bearings is very important. However, there are kinds of mechanical faults that occur frequently and cause great casualties and economical loss. In order to keep the locomotives performing at its best, different methods of fault diagnosis have been developed and used effectively to detect the machine faults at an early stage, such as the neural network, fuzzy theory. They can have an effective diagnosis to the different faults, but to the diagnosis of the different stages of one fault, the existing methods can't get good effect. And the diagnosis of the different stages of one fault is especially important and has great influence on the monitoring and diagnosis of machineries. In this paper, Granular Computing (GrC) is used. The basic idea of GrC is that a complicate problem can be divided into several small ones which can be easily understood and solved according to the idea of GrC. Therefore, by constructing granularity structure, the faults information may be decomposed into different granularity levels, and the information of each level can be clearly understood and analyzed. The relations of the granules and the different levels may provide us a good method to distinguish the

different faults. Generally, we construct the granularity structure using the features exacted from the vibration signals of the machineries. However, some of the features contain too much unrelated information to the faults and there is a high degree of overlap among the values of these features of different faults. These features would confuse the granular structure and therefore, cause great performance degradation. The effect of the granularity structure may be greatly different with different features. So it's very necessary to select the most useful features. Here, distance evaluation technique [10] is used to select the most superior features from the original features set.

Naturally, study of related models of granular computing is a meaningful direction. In the past ten years, there have been some studies about models of granular computing. Zadeh[4] proposed a general framework of granular computing based on fuzzy set theory. Lin [5, 6] and Yao[7] considered a model of granular computing using neighborhood systems. Pawlak[8] built a model of granular computing based on rough set theory. Recent years, shi zhongzhi[1] proposed a new mode called tolerance granularity space which has been proved to have good performance on information classification. In this paper, we propose a new approach to intelligent fault diagnosis of machinery based on tolerance granularity space.

The organization of this paper is as follows: we introduce the basic concept about tolerance granularity space in section 2. Faults diagnosis mode is constructed in Section 3. In Section 4, we verify the good performance of this approach by means of the data of rolling element bearings.

2 Classification Based on Tolerance Granularity Space

We assume that information about objects in a finite universe is given by a decision table $DT = (U, C \cup D, V_a, f)$. The details can be found in Ref. [14]. And let $B \subseteq C$. Then the rule set F generated from DT and B consists of all rules with the form as follows:

$$\bigwedge \{(a, v) : a \in B \text{ and } v \in (V_a \cup \{*\})\} \rightarrow d = v_d. \quad (1)$$

Where $v_d \in V_a$. The symbol “*” means that the value of the corresponding attribute is irrelevant to the rule. The length of the rule is the number of attribute whose values are not “*”. For example, in a decision system with 4 condition attributes (a_1, \dots, a_4) , $(a_1 = 1) \wedge (a_2 = *) \wedge (a_3 = 1) \wedge (a_4 = 1) \rightarrow d = 4$ is a rule, we describe the rule as a vector $(1, *, 1, 1, 4)$, and the length of the rule is 3.

2.1 Decision Rules and Granular Structure

The detailed description about tolerance granularity space can be found in paper [1] [11-12]. First, the object system is defined. The decision table DT [14] is composed by decision objects with the form: $object = (v_0, v_1, \dots, v_{n-1}, v_n)$, where v_i is the discretized eigenvalue and v_n is the classification label of this decision object. The decision table is just the training sample set. And then, we define the tolerance relation system. The tolerance proposition as follows:

$$cp(\alpha, \beta \mid DIS, D) \Leftrightarrow dis(\alpha, \beta \mid w) \leq d .$$

Where the distance function is:

$$\begin{aligned}
 dis(\alpha, \beta \mid w_j) &= \sum_{i=0}^n w_j (\alpha_i \oplus \beta_i), \\
 \alpha &= (\alpha_0, \alpha_1, \dots, \alpha_n), \\
 \beta &= (\beta_0, \beta_1, \dots, \beta_n), \quad \alpha \text{ and } \beta \in U \\
 \alpha_i \oplus \beta_i &= \begin{cases} 0 & |\alpha_i - \beta_i| \leq r \\ 1 & \text{else} \end{cases}
 \end{aligned} \tag{2}$$

$\alpha, \beta \in DT$, and for two objects, the distance function denotes the number of attributes, whose difference value of corresponding eigenvalues are less than or equal to r , among the front $n-1$ attributes.

$w = (w_0, w_1, \dots, w_n)$, $w_i=1$ or 0 , is the dimensional weight, which is an important parameter The tolerance relation can be adjusted dynamically by changing the w for getting different attributes set B of the two objects. And $w_n=0$, because the last dimension denotes the decision label of the objects, so it isn't involved in the calculation.

A tolerance granule (TG) is composed by two parts [13]: the intension of TG , named IG , which is the decision rule defined in formula (1); the extension of TG , named EG , which contains all the objects satisfying IG in the decision table. That is, $TG = (IG, EG)$. Therefore, if all the classification labels of the objects in EG are the same, we call the TG consistent tolerance granules (CTG), the IG of this TG can be a classification rule, or else not.

A granular structure have m levels, m is equal to the dimension of w . when only one dimension is "1" in w , others are "0", and let "1" ergodics every dimension in w , we can get all the 1th-level granules under different forms of w : $G_1 = \{TG_{1,1}, \dots, TG_{1,s}\}$. And each w corresponds with a granule. Then we get the j th-level granules by granulating every granule in the $(j-1)$ th-level by means of the formula as follow:

$$\begin{aligned}
 EG_j &= \{x \mid (x, \eta_j) \in cp(\alpha, \beta \mid DIS, D) \wedge x \in EG_{j-1}\} \\
 \eta_j &\in EG_{j-1}
 \end{aligned} \tag{3}$$

Where w_j can be got by ergodic every "0" with "1" in w_{j-1} , and remaining the "1" formerly in w_{j-1} . Until all the dimension of w are "1", the last level granules are got. And each of the higher level granules is a sub-granule of a certain granule in front level. A granule may have several sub-granules. After getting the EG of each TG in per level, and if the TG is CTG , the classification rule can be formed as follows:

$$IG_{i,j} = \eta_{i,j} * w_{i,j} \tag{4}$$

That means the values in IG corresponding with the "0" in w are "*", and the others are the values in η corresponding with the "1" in w . The length of the rule shows the level to which the TG belongs.

So, the more the attributes contains, the finer the objects set is granulated. We can select different levels for solving our problem according to different situation.

2.2 Classification

As an object for classifying, beginning from the first level, we search the consistent tolerance granules matching with the object using the classification rules. Here, we define the confidence and support to measure the rules. If the objects can't be distinguished in this level, we search in the next level. It contains more fault information than the above level.

3 Fault Diagnosis Based on Tolerance Granularity Space Mode

The main task of intelligent fault diagnosis is faults classification. We need distinguish the different faults as well as the different stages of a fault with a high accuracy. Here, the different faults and the different severities of one fault can be divided into different information granules, and each granule may contain the different fault or the

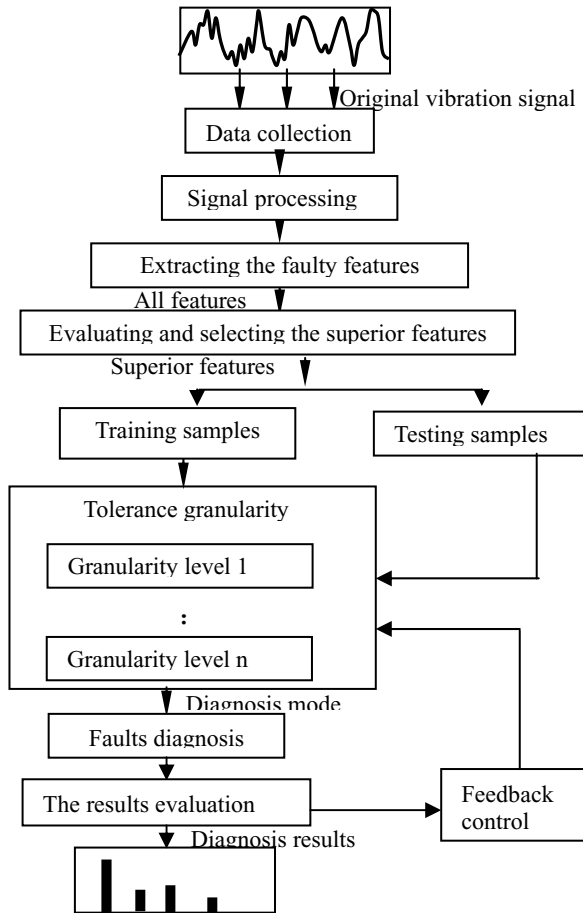


Fig. 1. The mode of intelligent fault diagnosis

information of the different stages of one fault by constructing the tolerance granularity space mode. Fig-1 shows the mode of intelligent fault diagnosis. We can know that in a decision table, $DT = \langle U, C \cup D, V_a, f \rangle$, all the sample signals consist of the training samples set U , C is a feature set. And D contains the different faults expressed by different number. V_a contains all the eigenvalues of the features.

Before a feature set is fed into this mode, most superior features providing dominant fault-related information should be selected from the whole feature set, and irrelevant or redundant features must be discarded to improve the classifying performance and avoid the curse of dimensionality. Here, distance evaluation technique [10] can be used to select the most superior features from the original features. The features with the smallest distance evaluation criterion α_j will be the most superior features. As shown in formula (5), where, $d_j^{(w)}$ is the average distance of the C faults. $d_j^{(b)}$ is the average distance between the different faults samples.

In formula (2), we set $d=0, r=DIS * w$. DIS is a $n-1$ dimension vector composed by dis_j , calculated as follows:

$$\begin{aligned}
 d_{c,j} &= \frac{1}{M_c \times (M_c - 1)} \sum_{l,m=1}^{M_c} |v_{m,c,j} - v_{l,c,j}|, \\
 l, m &= 1, 2, \dots, M_c, l \neq m \\
 d_j^{(w)} &= \frac{1}{C} \sum_{c=1}^C d_{c,j}, u_{c,j} = \frac{1}{M_c} \sum_{m=1}^{M_c} v_{m,c,j} \\
 d_j^{(b)} &= \frac{1}{C \times (C - 1)} \sum_{c,e=1}^C |u_{e,j} - u_{c,j}| \\
 \alpha_j &= \frac{d_j^{(w)}}{d_j^{(b)}} \\
 dis_j &= \min(d_{c,j}), \quad j = 1, 2, \dots, n - 1 \\
 DIS &= (dis_1, dis_2, \dots, dis_{n-1})
 \end{aligned}
 \tag{5}$$

Where, $v_{m,c,j}$ is the j th eigenvalue of the m th sample under the c th condition, M_c is the sample number of the c th condition. $d_{c,j}$ is the j th inner-class distance under the c th condition. C is the number of the conditions.

So, if the samples x and x_j satisfy the tolerance proposition, the tolerance granules TG are generated as:

$$\begin{aligned}
 EG &= (x, x_j) \Leftrightarrow dis(x, x_j | w_i) \leq 0 \\
 x &= (x_0, x_1, \dots, x_n) \\
 x_j &= (x'_0, x'_1, \dots, x'_n) \\
 x \oplus x_j &= \begin{cases} 0 & |x-x_j| \leq DIS * w_i \\ 1 & else \end{cases}
 \end{aligned}
 \tag{6}$$

Where x is replaced by the next samples after the x_j ergodic all the samples in the training set.

Here, in order to know whether the TG is CTG or not, we define the confidence of the granule TG [13]:

$$Conf(c_i)=confidence(class=c_i|a=v)=P(class=c_i | a=v)$$

$$P(class = c_i | a = v) = \frac{M_{class=c_i \wedge a=v}}{M_{a=v}} . \tag{7}$$

Where $M_{class=c_i \wedge a=v}$ is the number of the samples with condition $c_i \in D$ in the TG and $M_{a=v}$ is the number of the samples with attribute set $a \subset C, v \in V_a$. Because some features of the different faults may overlap with the others, this will affect the judgment to the TG . So, for every TG , we define a threshold value T_value as follows:

$$IG = \begin{cases} (r_1, r_2, \dots, r_{n-1}, c) & \max(conf) \geq T_value \\ 0 & else \end{cases} . \tag{8}$$

$$r_i = \frac{1}{M} \sum_{p=1}^M v_{i,j}, 0.5 \leq T_value \leq 1$$

Where T_value denotes the reliability of the rule and we can give it different values in different situation. r_i is the average of all the eigenvalues of each feature in the TG corresponding with the positions in $w=1$, and $r_i=0$ in $w_i=0$. c is the condition label with the maximum $Conf$ in each TG .

As shown in Fig-1, we extract the eigenvalues of the superior features of the testing samples by distance evaluation technique. And then we calculate R from the first level as follows:

$$R = \{IG_{i,j}, \dots\} \Leftrightarrow dis(IG_{i,j}, y | w_{i,j}) \leq 0 . \tag{9}$$

Where, i denotes the i th level, j denotes the j th granule in i th level.

Step 1: If all the IGs in R are the same c , the condition of y is c .

Step 2: If all the IGs in R are not the same c , and any granule has sub-granules, we calculate in the next level in the same way as step 1.

Step 3: if all the IGs in R are not the same c and none of the granules has sub-granules, we get the condition of y as follows:

$$Support_{i,j} = \frac{M_{EG_{i,j}}}{M_{TG}} . \tag{10}$$

$M_{EG_{i,j}}$ and M_{TG} are the number of samples in the $EG_{i,j}$ and in the TG , respectively. So the IG with the maximal of $Support$ decides the condition of y . But if the TG with the maximal $Support$ is not one, we calculate P as follows:

$$P_s = P(|y - R_s|^2) = \sum_{j=0}^{n-1} (y_j - r_j)^2, y = (y_0, \dots, y_{n-1}), \tag{11}$$

$$R_s = (r_0, \dots, r_{n-1},)$$

R_s is the $IG_{i,j}$ with the maximal support. Then, the IG with the minimal P decides the condition of y .

4 Experiments

As the large-scale complex electromechanical system, there are many kinds of parameters which must be monitored for locomotives. And the relations between these parameters are complicated, high-randomness. Meanwhile, there are many kinds of faults among which sudden faults and compound faults take up a great proportion. The wheel set bearings' faults are the typical ones, so there is great practical engineering value to have a correct recognition to the wheel set bearings' faults of locomotives. In this paper, we analyzed the wheel set bearings' faulty condition of locomotives based on tolerance granularity space mode, and the data were got from the wheel set bearings' experimental setup, the type is JL-501A.

Fig-2(a) shows the wheel set bearings' experimental setup. The motor drives the bearing to rotate, and the speed is about 500 rpm. We collect the vibration data from the testing bearing in the 1 and 2 measuring points by means of accelerometer. Fig-2(b) shows the schematic of the experimental setup. In present paper, the original data is divided into some samples with 4096 data points.

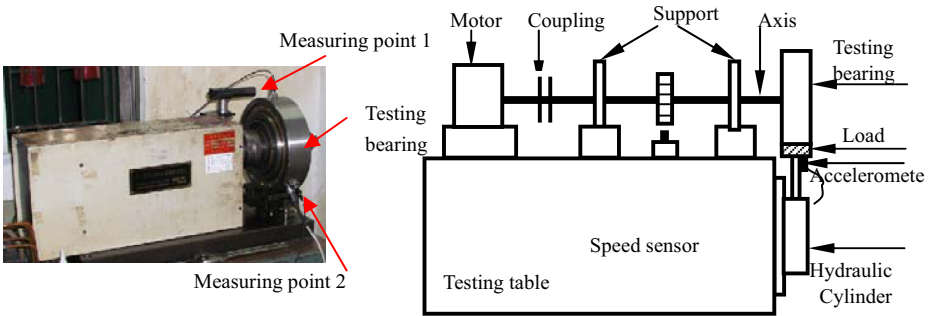


Fig. 2. (a) The experimental setup

(b) Schematic of the experimental setup

Table 1. Description of bearing data

Table 2. bearing conditions

Parameters	Value	label	Operating condition
Bearing type	552732QT	1	Normal
Load	9800N	2	Outer race-early fault
Inner-race diameter	160mm	3	Outer race-severe fault
Outer-race diameter	290mm	4	Inner race
Ball diameter	34mm	5	Ball
Ball number	17	6	Inner race and outer race-compound fault
Contact angle	0°	7	Outer race and ball compound fault
speed	480~640rpm	8	Inner race and ball compound fault
Sampling frequency	12.8kHz	9	Outer race, inner race and ball compound fault
Sampling length	8192		

The test parameters of the locomotive bearing in this experiment are shown in Table 1. The collecting data includes 9 subsets, and each subset corresponds with one faulty condition (including the normal condition), each kind of fault condition contains 80 samples. Table 2 shows the 9 kinds of fault conditions. Fig-3(a)-(d) shows the pictures of the different kinds of faulty bearings. And the 9 kinds of original vibration signal waveforms are shown in Fig-4.

The data set is composed by 720 samples with nine different operating conditions. And each fault condition has 40 samples for training, the rest 40 are testing samples.

For every sample, we decomposed it into eight frequency bands by means of the second generation wavelet technique, and extracted the 13 frequency domain features for each band. Then we got the feature set containing 117 features together with the 13 frequency domain features of the original signal. Fig-5 shows the distance evaluation criterion calculated by distance evaluation technique. Then we form the attribute set using these most superior features selected from the 117 features according to the distance evaluation criterion to construct the granular structure, respectively. Table 3 shows the classification accuracy with different number of superior features. Obviously, the highest accuracy is 91.94%.

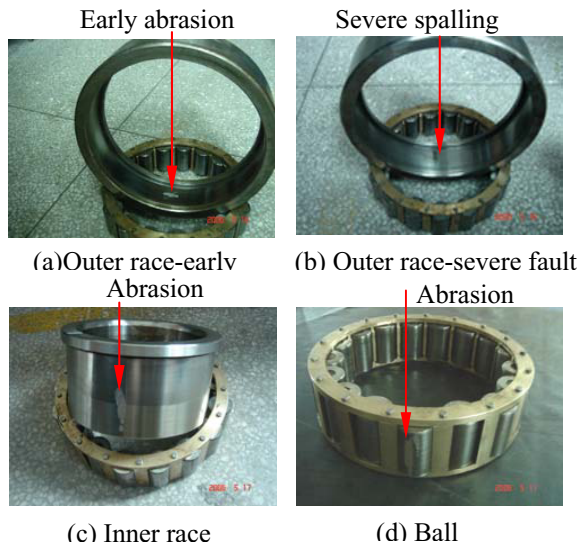


Fig. 3. The different kinds of fault bearings

Table 3. Performance comparison for different features

Feature number	Different superior features						6 time domain features
	Two	Three	Four	Five	Six	Seven	
Testing	90	90.28	90.57	91.01	91.94	90.57	71.11

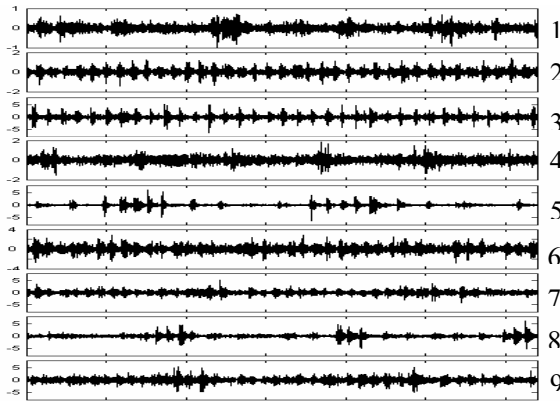


Fig. 4. Original time waveform

Here, we set the $T_value=0.9$. Obviously, the classification accuracy is the highest when selecting six superior features. However, when we select the six time domain features of the original signal, the rate of wrong classification is 28.89%, compared with 8.06% (six most superior features), it is much higher. This means that the finer we granulate, the more information about the faults we get. Meanwhile, the accuracy is decreased to 90.57% when we select seven superior features. This implies that the number of the superior features also has influence on the granularity structure, so it needs further study of the method that how to get the suitable features set to construct the granularity structure.

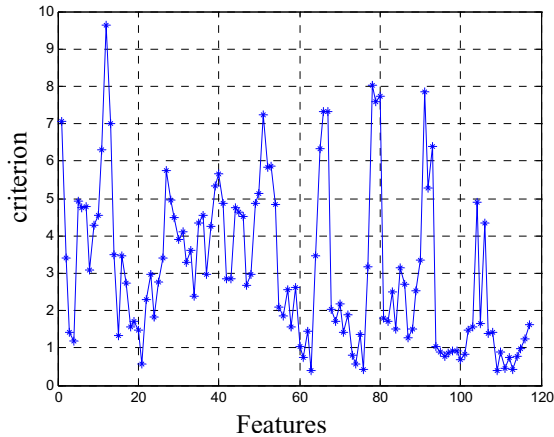


Fig. 5. Distance evaluation criterion of all features

This suggests that some of the features contain too much fault-unrelated information and there is a high degree of overlap among the values of these features under the nine fault conditions. These features can confuse the granular structure. Therefore, this causes great performance degradation. This implies that time-domain features provide too little bearing fault-related information, and therefore, are unable to distinguish the nine fault conditions.

As the diagnosis of the locomotive bearings, it needs to distinguish not only the fault categories including the outer race, inner race and ball faults, but also the early faults. That means we need distinguish both the slight abrasion fault and the severe spalling fault of outer race of the bearing shown in Fig-3. And especially important, the diagnosis of the compound faults is also very necces. This certainly makes the diagnosis to the bearing faults difficult. However, we can see from the experimental result that there is a high effectiveness of the intelligent fault diagnosis mode which integrates the second generation wavelet technique, distance evaluation technique and tolerance granularity space theory. Meanwhile, it denotes that we can get more fault-related information by decomposing the original signal by means of the second generation wavelet technique. This greatly improves the classification accuracy and the accuracy of the mode with different number of features is all more than 90%.

5 Conclusion

In this paper, we proposed a new approach to the intelligent fault diagnosis of mechanical equipments based on tolerance granular space and applied it to the condition recognition of locomotive bearings. We selected two to six most superior features as attributes respectively, and get the high classification accuracy 91.94% under six features. The experimental results show that the construction of the attribute space has an important influence to the performance of the granular structure. And the approach enables the diagnosis of abnormalities in locomotive bearings and at the same time identification of the categories and severities of faults with a high accuracy. This approach not only extends the application field of granular computing (GrC), but also introduces a new and effective method for fault diagnosis.

Acknowledgement. This paper is supported by the Natural Science Foundation of China “Research on Hybrid Intelligent Technique and Its Application in Fault Diagnosis Based on Granular Computing” (Grant No.50875197) and the Project Sponsored Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry.

References

- [1] Zheng, Z., Hu, H., Shi, Z.-Z.: Tolerance Relation Based Granular Space. In: Ślęzak, D., Wang, G., Szczuka, M.S., Düntsch, I., Yao, Y. (eds.) RSFDGrC 2005. LNCS, vol. 3641, pp. 682–691. Springer, Heidelberg (2005)
- [2] Liu, B., Ling, S.F., Gribonval, R.: Bearing Failure Detection Using Matching Pursuit. *NDT & E International* 35, 255–262 (2002)

- [3] Nikolaou, N.G., Antoniadis, I.A.: Rolling Element Bearing Fault Diagnosis Using Wavelet Packets. *NDT & E International* 35, 197–205 (2002)
- [4] Zadeh, L.A.: Towards A Theory of Fuzzy Information Granulation and Its Centrality in Human Reasoning and Fuzzy logic. *Fuzzy Sets and Systems* 19, 111–127 (1997)
- [5] Lin, T.Y.: Granular Computing on Binary Relations(I). In: Polkowski, L., Skowron, A. (eds.) *Rough Sets in Knowledge Discovery. Methodology and Applications*, vol. 1, ch. 6, pp. 107–121. Physica-Verlag, Heidelberg (1998)
- [6] Lin, T.Y.: Granular computing on binary relations(II). In: Polkowski, L., Skowron, A. (eds.) *Rough Sets in Knowledge Discovery. Methodology and Applications*, vol. 1, ch. 7, pp. 122–140. Physica-Verlag, Heidelberg (1998)
- [7] Yao, Y.Y.: Relational Interpretations of Neighborhood Operators and Rough Set Approximation Operators. *Information Sciences* 111, 239–259 (1998)
- [8] Pawlak, Z.: Granularity of Knowledge. Indiscernibility and rough sets, 106–110 (1998)
- [9] Pei, D.W.: Some Models of Granular Computing. In: 2007 IEEE International Conference on Granular Computing, pp. 17–22 (2007)
- [10] Lei, Y., et al.: A New Approach to Intelligent Fault Diagnosis of Rotating Machinery. *Expert Systems with Applications* (2007), doi:10.1016/j.eswa.2007.08.072
- [11] Zheng, Z., Hu, H., Shi, Z.Z.: Tolerance Granular Space and Its Applications. In: IEEE International Conference on Granular Computing, pp. 367–372 (2005)
- [12] Zheng, Z., Hu, H., Shi, Z.Z.: Granule Sets Based Bilevel Decision Model. In: Wang, G.-Y., Peters, J.F., Skowron, A., Yao, Y. (eds.) *RSKT 2006. LNCS (LNAI)*, vol. 4062, pp. 530–537. Springer, Heidelberg (2006)
- [13] Yao, J.T., Yao, Y.Y.: A Granular Computing Approach to Machine Learning. In: *Proceedings of the 1st International Conference on Fuzzy Systems and Knowledge Discovery*, pp. 732–736 (2002)
- [14] Yao, Y.Y.: Potential Applications of Granular Computing in Knowledge Discovery and Data Mining. In: *Proceedings of World Multi-conference on Systemics, Cybernetics and Informatics*, pp. 573–580 (1999)

Analysis of Two Neural Networks in the Intelligent Faults Diagnosis of Metallurgic Fan Machinery

Jiangang Yi¹ and Peng Zeng²

¹ College of Machinery & Automation, Wuhan University of Science and Technology,
Wuhan 430081, China

² College of Mathematics & Computer Science, Jiangnan University,
Wuhan 430056, China

Yjg_wh@yeah.net, ZENG-PENG@sohu.com

Abstract. With the aim to study the faults diagnosis ability of the BPNN and the RBFNN, many experiments are done to test the learning ability, the diagnosis ability and the anti-noise ability. The analysis shows the RBFNN has better learning ability and anti-noise ability than the BPNN. However, in the process of concurrent faults diagnosis, both have bad recognition rate. A realistic application verifies the single neural network can not used for metallurgic fan machinery faults diagnosis.

Keywords: ANN, Metallurgic fan machinery, Intelligent faults diagnosis.

1 Introduction

In the process of mechanical faults diagnosis, it is a complex nonlinear inference process to judge the faults of the objects with the acquired data. Because artificial neural network (ANN) can solve the problem of learning the map between faults and signals without pre-knowledge and functions, it is widely used in faults pattern recognition [1,2]. Metallurgic fan machinery includes the air ejector fan, air blaster, dedusting fan and so on which are used in the smelting process. For worked in bad environment of high temperature, high pressure and high dust content, metallurgic fan machinery is out of work easily. Therefore, to realize the intelligent faults diagnosis of metallurgic fan machinery with ANN is very important and valuable.

Back propagation neural network (BPNN) and radial basis function neural network (RBFNN) are two supervised algorithms which are widely used in these years. BPNN is a multi-layer forward spread network with min mean square deviation learning method; RBFNN is a partly approximate network with radial basis function in hidden layer. To analyze the faults diagnosis effects of two neural networks in the application of metallurgic fan machinery, three aspects of network learning ability, network diagnosis ability and network anti-noise ability are done with the collected data with the rotor experiment table shown in Fig.1.

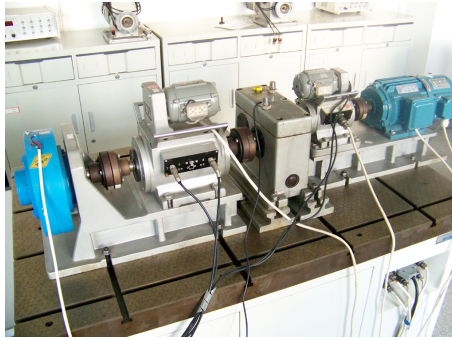


Fig. 1. Rotor experiment table

2 Sample Sets Building

The vibration signal energy of different frequency in frequency domain is usually used as faults characteristic during the faults diagnosis based on ANN. It means when the monitoring signal from the vibration sensor in some position is above the preset alarming value, the collected time wave in this time should be transferred to frequency wave with FFT algorithm, the amplitude value in the specific position should be extracted as the characteristic value, which can be used as input samples after normalization. The faults sort can be acquired with suitable ANN algorithm and the input sample.

In order to identify the main faults of metallurgic fan machinery such as unbalanced rotor, noncentering rotor and vortex oil film and so on, nine peak frequencies in the position of $(0.01\sim 0.49)f$, $0.50f$, $(0.51\sim 0.99)f$, $1f$, $2f$, $3f$, $4f$, $5f$ and $>5f$ in frequency domain are used as the characteristics (f means line frequency). In the rotor experiment table, many experiments are tested with the characteristics to collect the data. The results can be used as the input-output sample sets of neural network. In this field professor Yu Hejie in china has done much [3].

3 Neural Network Learning Ability

The thesis has proved the neural network with single hidden layer can map all nonlinear function [4], so a BPNN and a RBFNN composed with an input layer, a hidden layer and an output layer are designed to test their ability in the process of faults diagnosis. In the experiment, the Matlab neural network tool set is used to build the BPNN (momentum coefficient is 0.5 and learning speed rate self adaptive adjusted) and the RBFNN. Fig.2 shows the parameter setting interface of the BPNN in Matlab [5].

The learning ability of neural network is usually judged by mean-squared error (MSE). In Fig.3, when the training MSE goal is 0.01, the condition of convergence is reached after 5000 training times with the BPNN. However, in the same condition, only 350 training times are needed to reach the same target with the RBFNN in Fig.4. This indicates the RBFNN faults diagnosis time is shorter and the real-time diagnosis ability is better compared with the BPNN.

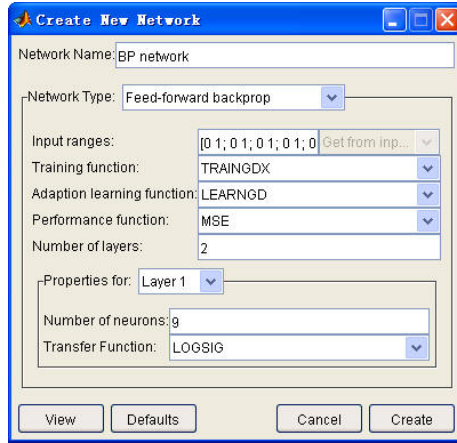


Fig. 2. Parameter setting interface

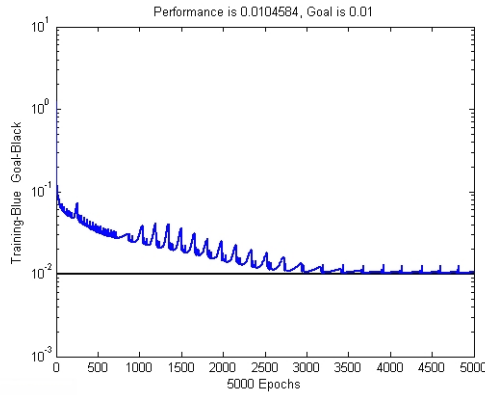


Fig. 3. The BPNN learning ability

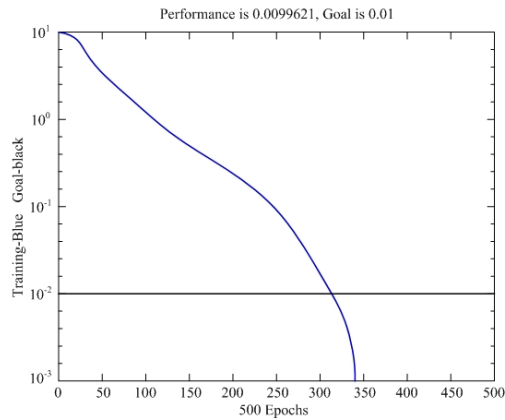


Fig. 4. The RBFNN learning ability

4 Neural Network Diagnosis Ability

To test the diagnosis ability of the BPNN and the RBFNN, four types of single faults experiments of unbalanced rotor (A), noncentering rotor (B), active cell and stator friction (C), and supporting parts loosening (D) have been done for 50 times, the results are shown in Table 1. In Table 1, the diagnosis accuracy rate of all single faults is above 90%, which indicates both the BPNN and the RBFNN have good diagnosis ability for single faults.

Table 1. The diagnosis accuracy rate

Faults Sort	Algorithm	Fault Number	Leak Number	Accuracy Rate
A	BPNN	4		92%
	RBFNN	3		94%
B	BPNN	4		92%
	RBFNN	5		90%
C	BPNN	5		90%
	RBFNN	4		92%
D	BPNN	5		90%
	RBFNN	5		90%
A+B	BPNN	8	7	70%
	RBFNN	7	7	72%
A+C	BPNN	10	10	60%
	RBFNN	9	7	68%
A+B+C	BPNN	10	13	54%
	RBFNN	9	11	60%

For the purpose of testing the diagnosis ability of concurrent faults of the BPNN and the RBFNN, the experiments of A+B, A+C and A+B+C have been done and the results are in Table 1. It is shown in Table 1 that neither the BPNN nor the RBFNN has good diagnosis ability when the concurrent faults samples are used as neural network input data, which indicates the two algorithms can not be applied to diagnose concurrent faults directly.

5 Neural Network Anti-noise Ability

Because the input signals and the training sample sets often contain noise, the training samples without noise and with 20% noise level are used to train the BPNN and the RBFNN. Then the test samples with 0% to 25% noise level are used to verify the anti-noise ability. The recognition abilities are shown in Fig.5 and Fig.6. Fig.5 shows the BPNN with noise and the BPNN without noise have similar anti-noise abilities, which

indicates the BPNN has good adaptability for the training sample with some noise. When the noise level is less than 20%, the error recognition rates of both networks are less than 15%. But when the noise level is above 20%, the error recognition rates increase evidently. Compared with Fig.5 and Fig.6, the error identification rate curves of RBFNN with and without noise change gently even when the noise level is up to 25%. This presents the RBFNN has better anti-noise ability than the BPNN.

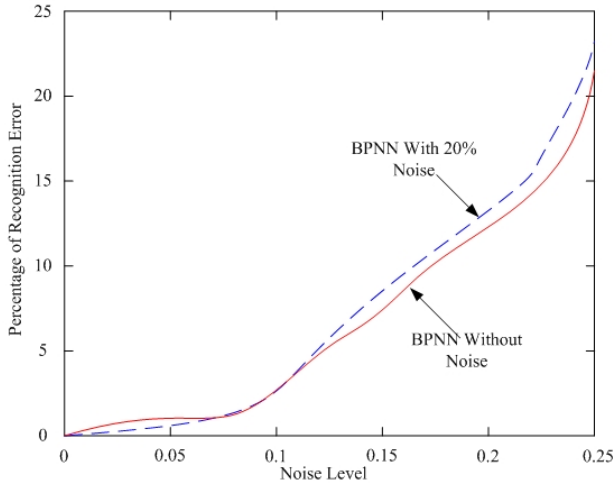


Fig. 5. The identification ability of the BPNN

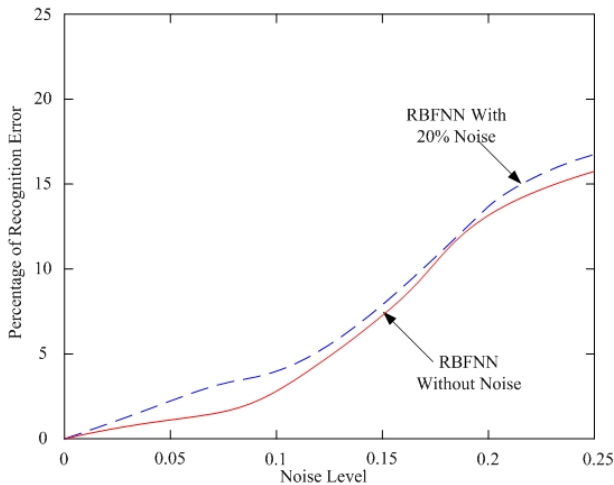


Fig. 6. The identification ability of the RBFNN

6 Application

The main rotor rotational speed of a metallurgic fan is 598RPM. The vibration value in the vertical direction of load side is 9.07mm/s in some time, which exceeds the preset alarming value. The characteristic values on the horizontal and the vertical direction of the load sides and the unload sides in the time are recorded as Table 2 shows.

Table 2. The vibration values of the metallurgic fan

Frequency Channel	(0.01 to 0.49)f	0.5f	(0.51 to 0.99)f	1f	2f	3f	4f	5f	>5f
Horizontal direction of load side	0.02	0.00	0.00	0.90	0.45	0.00	0.08	0.00	0.10
Vertical direction of load side	0.00	0.00	0.03	0.85	0.44	0.23	0.12	0.00	0.02
Horizontal direction of unload side	0.00	0.02	0.00	0.54	0.49	0.23	0.05	0.00	0.02
Vertical direction of unload side	0.00	0.01	0.00	0.83	0.36	0.17	0.15	0.15	0.12

The diagnosis conclusion of the BPNN with the sample in Table 2 is rotor unbalancing, while the conclusion of the RBFNN is rotor unbalancing and supporting parts loosening. After the maintenance, it is found the fault of this time is rotor unbalancing and supporting parts loosening and active cell and stator friction, which indicates both the BPNN and the RBFNN have missed diagnosis.



Fig. 7. The Metallurgic Fan Rotor Faults Diagnosis

7 Conclusion

From the experiments data and the application data, it is shown the RBFNN has better learning ability and anti-noise ability than the BPNN. However, when judged from diagnosis ability, neither the RBFNN nor the BPNN has good diagnosis accuracy rate for concurrent faults. This is because the input signals are from only one sensor, thus both neural networks memorize only part character of the whole system, and can not precisely classify complex faults. To solve this problem, only change neural network algorithm is not enough for metallurgic fan machinery faults diagnosis.

Acknowledgements. This work is partially supported by the Natural Science Foundation of Hubei province granted 2008ABA003.

References

1. Zhu, D., Yu, S.: Neural Networks Data Fusion Algorithm of Electronic Equipment Fault Diagnosis. In: The 5th World Congress on Intelligent Control and Automation, pp. 14–18 (2004)
2. Su, L., Goh, et al.: A Complex Valued RTRL Algorithm for Recurrent NN. *Neural Computation*, pp. 2699–2713 (2004)
3. Yu, H., Han, Q., et al.: *Devices Faults Diagnosis Engineering*. Metallurgical Industry Press, Beijing (2001)
4. Vladimir, N.V.: *Nature of Statistics Learning Theory*. Springer, New York (2000)
5. The MathWorks Inc. (2004), <http://www.mathworks.com>
6. Islam, M., et al.: A Construction Algorithm for Training Cooperative NN Ensembles. *IEEE Trans. NN*, 820–834 (2003)

Research on the Diagnosis of Insulator Operating State Based on Improved ANFIS Networks

Zipeng Zhang¹, Shuqing Wang¹, Liqin Xue², and Xiaohui Yuan³

¹ Hubei University of Technology, Wuhan, 430068, China

² Shandong University of Science and Technology, Qingdao, 266510, China

³ Huazhong University of Science and Technology, Wuhan, 430074, China
zhzip1@163.com

Abstract. Power transmission line insulator is an important part for power system security. Because insulator has complex operating environment and its infection factors interact on each other, the diagnosis of insulator running state is very difficult. It is needed to use some useful information to conclude insulator operating state. ANFIS networks have strong knowledge expressing, fuzzy reasoning and learning ability, which is used to diagnose insulator operating state in this paper. In order to improve the knowledge gain ability for ANFIS, two ways both neural networks learning and GAs optimizing are combined together to train optimized reasoning parameters in reasoning rules training. The new improved quick GA is designed to train parameters based on the character of inference system. Experiment results show that the designed ANFIS network has strong reasoning and learning ability, which can diagnose insulator operating state unfaillingly. These techniques used in this paper are quite effective in expert diagnosis system.

Keywords: Insulator operating state, Diagnosis, ANFIS network, GAs, Improved learning way.

1 Introduction

Power transmission line insulator has complex operating environment in outdoors. It endures not only mechanical pressure, but also the impact of climate and environment. The sedimentary contamination of insulator's surface would make insulation ability decline when insulator is laid in dankish environment chronically. In the long run, foul contamination will cause insulator deterioration, flashover and so on [1],[2]. If it is not dealt with timely, power system reliability would be affected seriously [3],[4].

The diagnosis of insulator' running state in power line is main to forecast and diagnose insulator flashover, contamination degree, deterioration degree of insulator and conclude failure sources [5]. Because of insulator having complex operating environment and its infection factors having interaction, it is very difficult to diagnose insulator running state in power system. Now, it is an important task for the domain of power transmission to research the diagnosis and prediction system of the running

state for insulator [6]. In this paper, the expert diagnostic system for the contamination degree of insulator is designed to diagnose insulator running state [7]. In the diagnostic system, the contamination degree of insulator is diagnosed according to the metrical leakage current, pulse frequency, the equivalent humidity environment. The designed reasoning method solves the difficult diagnostic problem for insulator running state successfully.

2 Diagnostic System for Insulator

It is difficult to directly monitor the contamination degree of insulator. Here, an indirect measurement method is used to diagnose insulator running state. In the diagnosing, the easy measured leakage current, pulse frequency, the equivalent environment humidity are used to analyse and diagnose insulator running state[8]. Those signals are measured with on-line mode and then are used to analysis with off-line mode. The correlative data may be analysed or processed and display through graphics in the back computer.

According to national standard, the monitor content of air quality index includes salt contamination and insoluble ash contamination. In different regions or different years or different components, the effect of ESDD is different to insulator contamination [9],[10].

The SO_2 , NO_x and TSP(total suspended particulates) are main pollution sources. The SO_2 of the atmosphere and its complex sulphate are main salt pollutant in the insulator surface. The content parameter of SO_2 is used as main pollutants parameter and NO_x is used as assistant pollutant parameter. The Q_s is expressed as salt density index of SO_2 and NO_x . Q_N is expressed as ash density index for TSP.

$$\begin{cases} Q_s = \sqrt{\frac{M_{SO_2}}{B_{SO_2}}} \frac{1}{2} \left(\frac{M_{SO_2}}{B_{SO_2}} + \frac{M_{NO_x}}{B_{NO_x}} \right) \\ Q_N = \frac{M_{TSP}}{B_{TSP}} \end{cases} \quad (1)$$

Here, M is real measured density of pollutant and B is evaluating standard.

The line combination both salt density index Q_s and ash density index Q_N are contamination degree. The combining expression is given as follow.

$$\rho = Q_s \omega_1 + Q_N \omega_2 \quad (2)$$

Here, ω_1 and ω_2 are coefficient, which is decided according to region difference.

Contamination degree can be obtained through ANFIS net training, learning and reasoning. Because temperature change brings little effect, its effect is not calculated. The leakage current, pulse frequency and equivalent environment humidity are used as input for ANFIS net.

3 Design of ANFIS Network

3.1 The Structure of ANFIS Network

The designed ANFIS network has Takagi-Sugeno fuzzy reasoning rules. The net includes two parts. The first part net is used to match the fuzzy rules and the second part network is used to generate fuzzy conclusions [11]. The structure is shown in Fig.1. All the network is divided into six layers, the first part includes four pieces and the second part includes two pieces [12].

The first layer: Nodes at layer 1 are input nodes that represent input linguistic variables. In this layer, each node is adjusted to fuzzy reasoning region. Linking coefficients are transformed based on input range [8].

$$O^1 = x'_k = \omega^1 \cdot x_k, \quad k = 1,2,3 \tag{3}$$

Nodes at layer 2 are term nodes that act as membership functions (MF) to represent the terms of the respective linguistic variable. With the use of Gaussian MF, the operations performed in this layer are given as follow:

$$\mu_k^l(x_k) = \exp\left(-\frac{(x'_k - c_{kl})^2}{\sigma_{kl}^2}\right), \quad l = 1,2, \dots, m_k \tag{4}$$

Where c_{kl} and σ_{kl} are center and width of the Gaussian MF respectively. m_k is the number of fuzzy segmentation for x_k . The total number of nodes is $M = \sum_{k=1}^N m_k$, N is input number. The output of this layer is given as follow.

$$O_i^2 = \mu_k^l, \quad l = 1,2 \dots m_k, k = 1,2 \dots N, i = 1,2, \dots M \tag{5}$$

The third layer: Each node at layer 3 is a rule node, which represents one fuzzy logic rule. Thus all nodes of this layer form a fuzzy rule base. Hence the rule nodes perform the fuzzy AND operation. The functions are given as follow:

$$O_i^3 = \omega_i = \min\{\mu_1^l, \mu_2^l, \dots, \mu_N^l\}, \quad l = 1,2 \dots m_k, i = 1,2 \dots M \tag{6}$$

The fourth layer: The node fitness is integrated.

$$O_i^4 = \varpi_i = \frac{\omega_i}{\sum_{i=1}^M \omega_i}, \quad i = 1,2 \dots M \tag{7}$$

The fifth layer: the transfer function for each node is linear function which expresses local linear model. The output of each adaptive node is given as follow.

$$O_i^5 = \overline{\omega}_i \cdot f_i = \overline{\omega}_i(p_{i0} + p_{i1}x_1 + p_{i2}x_2 + p_{i3}x_3), \quad i = 1,2 \dots M \tag{8}$$

Here, p_{ki} is a parameter for conclusion, $k = 0,1,2,3$.

The sixth layer: The total output is computed in the layer.

$$y = O^6 = \sum_{i=1}^M O_i^5, \quad i = 1, \dots, M \tag{9}$$

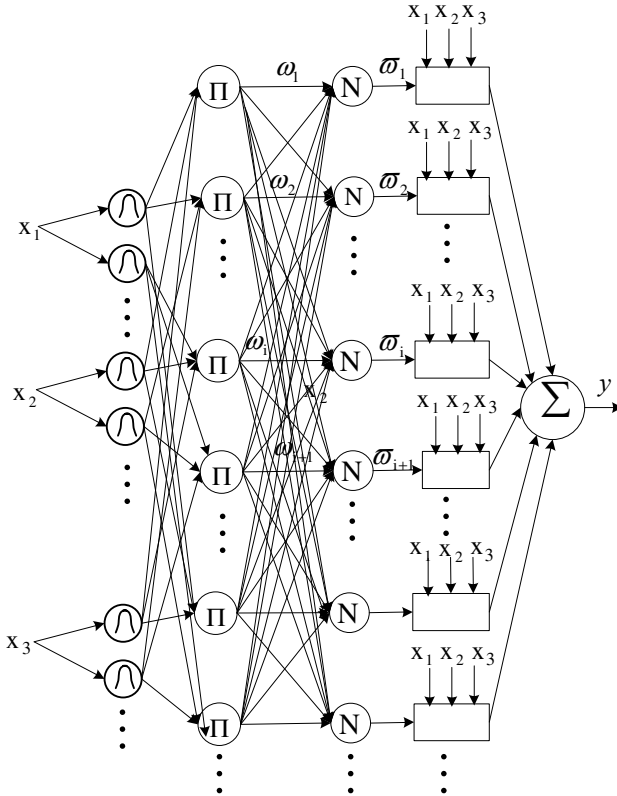


Fig. 1. Construction of ANFIS

3.2 Ascertaining of Reasoning Parameter

Gaussian MF parameters and reasoning parameters p_{ik} can be train through net learning. Because ANFIS net compartmentalizes two parts: anterior part and back part, all parameters cannot learn at the same time. When parameters of one part learn, another part parameters should be fixed. Here, back par parameters p_{ik} are ascertained through GA. And then Gaussian MF parameters are ascertained through neural network learning algorithm.

3.2.1 Optimizing of Genetic Algorithm

The designed GA has some amelioration based on general GA. The optimizing process of GA includes Encoding, Setting of initial population, Fitness calculating, Parent selection, Genetic operations and Judgement.

Before optimizing, some GA parameters need to set, such as popsize, lchrom, crossover probability P_c , mutation probability P_m , iterative number Gen, maxgen, maxruns.

In the optimizing of fuzzy inference parameters, general GA has some problems such as slow convergence speed, a great deal calculating, long identifying time and local optimization. In order to overcome those disadvantages, the new improved quick GA is designed based on the character of inference system. The improved arithmetic combines much strongpoint of other arithmetic, which can give best outcome or near best outcome. The improving of GA is given as follows:

1) *Encoding*

Because problem related to information is encoded in a structure called chromosome or string, the ranges of variation for the different variables should be selected through a careful study. In the training of p_{ik} , Supposing the fuzzy segmentation number of inputs is confirmed beforehand. Gaussian MF parameters are fixed according to even MF. Three factors need be trained. Here, eight bits binary code is adopted.

Where the bin (i) code of p_{ik} is shown as:

$$G(bin(i)) = (b_7 b_6 \dots b_0)_2 = (\sum_{i=0}^7 b_i * 2^i)_{10} \tag{10}$$

So every individual of scale parameters is formed by 24 bits binary gene code of chromosome.

2) *Setting of aim function*

The motive of GA optimizing is to make ANFIS reasoning net has appropriate parameters when input has big changes. The aim function can be expressed such as:

$$J(e) = \frac{1}{2} (\bar{y}_l - y_l)^2 \tag{11}$$

$$\begin{cases} F(e) = C_{max} - J(e), & J(e) < C_{max} \\ F(e) = 0 & , J(e) \geq C_{max} \end{cases} \tag{12}$$

There, C_{max} is appropriate multiple of pile value $J(e)$.

3) *Parent selection*

Parent selection mimics the survival of the best ones in the nature choice. At first, every individual's fitness and proportions of every individual's fitness are calculated. Then, selecting probability of every individual are decided according to the percent of individual's fitness. Here, the simulated annealing genetic algorithm is used here to draw properly fitness. The material drawing way is given as follow.

$$f_i = \frac{e^{f_i / T}}{\sum_{i=1}^M e^{f_i / T}} \tag{13}$$

$$T = T_0 (0.99^{g-1}) \tag{14}$$

There, f_i is the fitness of i th individual. M is popsize, g is iterative number Gen, T is temperature, T_0 is initial temperature.

Through drawing fitness continually using above way, some individuals, which have similar fitness, have like probability to be selected in early high temperature. With temperature dropping and fitness' being draw, the individuals' difference is magnified, which makes excellent individual has more obvious advantages.

4) Genetic operations

After selection of individuals according to their fitness values and gathering of selected individuals into a gene pool. Crossover is achieved in three stages: The first stage is matching. In the second stage, a crossover point is determined in each of the individuals. In the final stage, two parts of the individuals are replaced with each other. A probability test determines whether a mutation will be carried out or not.

3.2.2 Learning of Network

The center c_{kl} , width σ_{kl} of the Gaussian MF need to learn [11].

The supervised gradient decent method is used as learning algorithm for the designed ANFIS. The learning algorithm is given as follows. The error function E is defined as follow.

$$E = \frac{1}{2} (y_j - \bar{y}_j)^2 \tag{15}$$

The learning process c_{kl} and σ_{kl} is given as follow. Here p_{ik}^j is fixed.

$$\delta_j^6 = -\frac{\partial E}{\partial O_j^6} = -\frac{\partial E}{\partial y_j} = -(y_j - \bar{y}_j), \quad j=1,2 \dots r \tag{16}$$

$$\delta_j^5 = \delta_j^6 \tag{17}$$

$$\delta_i^4 = \frac{\partial E}{\partial O_i^4} = \frac{\partial E}{\partial y_j} \cdot \frac{\partial y_j}{\partial O_i^4} \cdot \frac{\partial O_i^4}{\partial O_i^5} = \sum_{j=1}^r \delta_j^5 f_i, \quad j=1,2 \dots r, i=1,2 \dots M \tag{18}$$

$$\delta_i^3 = \frac{\partial E}{\partial O_i^3} = \frac{\partial E}{\partial y_j} \cdot \frac{\partial y_j}{\partial O_i^3} \cdot \frac{\partial O_i^3}{\partial O_i^4} \cdot \frac{\partial O_i^4}{\partial O_i^5} = \delta_i^4 \cdot \frac{\sum_{x=1}^M \omega_{xi}}{\left(\sum_{x=1}^M \omega_{xi}\right)^2}, \quad x=1,2 \dots M, i=1,2 \dots M \tag{19}$$

$$\delta_{kl}^2 = -\frac{\partial E}{\partial O_i^2} = -\frac{\partial E}{\partial y_j} \cdot \frac{\partial y_j}{\partial O_i^2} \cdot \frac{\partial O_i^2}{\partial O_i^3} \cdot \frac{\partial O_i^3}{\partial O_i^4} \cdot \frac{\partial O_i^4}{\partial O_i^5} \cdot \frac{\partial O_i^5}{\partial O_i^6} \tag{20}$$

$$= \sum_{m=1}^M \delta_m^3 s_{kl} \cdot \exp\left(-\frac{(x_k - c_{kl})^2}{\sigma_{kl}^2}\right), \quad j=1,2 \dots r, l=1,2 \dots m_k$$

When μ_k^l is the minimal value of node m , formula (21) comes into reality.

$$s_{kl} = \frac{\partial O_i^3}{\partial \mu_k^l} = 1 \tag{21}$$

Otherwise, formula (22) comes into reality.

$$s_{kl} = \frac{\partial O_i^3}{\partial \mu_k^l} = 0 \tag{22}$$

The learning algorithm of c_{kl} and σ_{kl} is given as follow. β is given learning speed.

$$\frac{\partial E}{\partial c_{kl}} = -\delta_{kl}^2 \frac{2(x_k' - c_{kl})}{\delta_{kl}^2} \tag{23}$$

$$\frac{\partial E}{\partial \sigma_{kl}} = -\delta_{kl}^2 \frac{2(x_k' - c_{kl})^2}{\delta_{kl}^3} \tag{24}$$

$$c_{kl}(n+1) = c_{kl}(n) - \beta \frac{\partial E}{\partial c_{kl}} \tag{25}$$

$$\sigma_{kl}(n+1) = \sigma_{kl}(n) - \beta \frac{\partial E}{\partial \sigma_{kl}} \tag{26}$$

4 Analysis of Diagnosis Result

In this experiment, 100 groups experimental data is collected as training data and collected 10 groups experimental data is used as verifying data. After GAs and learning algorithm, the gained ANFIS is used to reason insulator contamination degree. The contrast between the ANFIS network output and the real data is given in fig.2. Here, sign * represents real measured data and sign □ represents the output outcome of the ANFIS network. The designed ANFIS has better learning ability and the trained error less than 0.00015 through 70 times training, which shows that the designed ANFIS has better convergence ability. The trained Gaussian MFs are given in fig.3, fig.4 and fig.5.

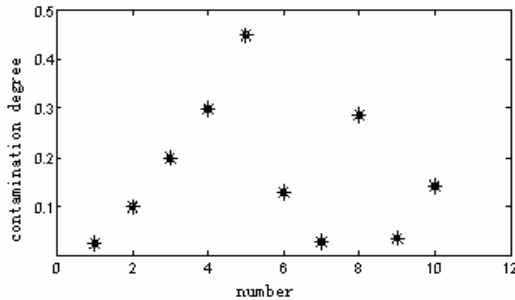


Fig. 2. Contrast between the output of ANFIS and real data

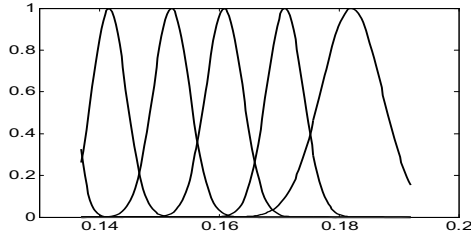


Fig. 3. Gaussian MFs of humidity input after training

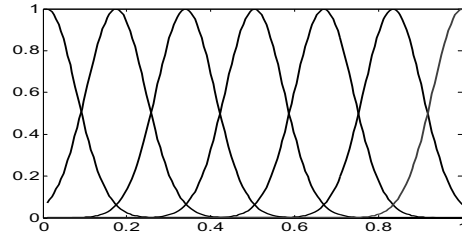


Fig. 4. Gaussian MFs of leakage current after training

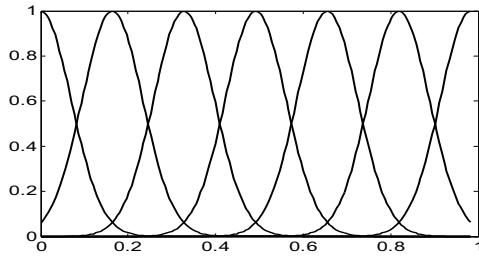


Fig. 5. Gaussian MFs of pulse frequency after training

Table 1. Diagnosis analysis of contamination instance

voltage grade /kV	environment humidity %	environment temperature	leakage current /mA	impulse number	contamination mg /cm ²	ANFIS output
35	78.0	19.0	0.32	0	0.025	0.0251
35	96.0	21.8	5.08	0	0.10	0.0999
35	90.2	18.0	15.54	56	0.2	0.1998
35	93.2	17.8	17.8	120	0.30	0.3000
35	89.5	19.0	56.03	165	0.45	0.4501
35	70.0	18.0	0.15	0	0.13	0.1295
35	70.4	18.7	0.29	0	0.028	0.0280
110	76.3	21.8	16.97	120	0.287	0.2876
110	68.5	23.7	2.38	0	0.035	0.0350
110	88.0	19.0	4.03	0	0.142	0.1423

The analysis for diagnosed outcome is given in table 1.

Here, impulse frequency is taken count of those impulses whose peak value is bigger than 300mA in one minute.

As seen from the table1, the diagnosis result through ANFIS network is consistent with the actual contamination and error is less than 0.0005 mg/cm². It is more direct and precise than general fuzzy diagnosis.

5 Conclusion

Insulator has complex operating environment and its infection factors have interaction. It is very difficult to diagnose the running state of insulator by using general way. ANFIS has better knowledge expression ability and learning ability for design dynamic knowledge expert system. Expert diagnosis inference system is designed for the diagnosis of insulator operating state based on ANFIS in this paper. In the diagnostic system, the contamination degree of insulator is diagnosed according to the metrical leakage current, pulse frequency, the equivalent humidity environment. The structure and reasoning process of ANFIS is reasonable. In order to gain optimization reasoning knowledge, GA is used to train rule parameters. The improved GA can gain preferable reasoning parameters through training. The experimentation result shows that the designed ANFIS has strong learning ability, which can diagnose insulator operating state reliably.

Acknowledgements. The authors gratefully acknowledge the financial supports from National Natural Science Foundation of China under Grant Nos. 50539140 & 50779020.

References

1. Riquel, G., Spangenberg, E.: Review of in Service Diagnostic Testing of Composite Insulators. *Electra* 169(12), 105–119 (1996)
2. Vlastos, A.E., Orbeck, T.: Out Door Leakage Current Monitoring of Silicone Composite Insulators in Coastal Service Conditions. *IEEE Transon Power Delivery* 11(2), 1066–1070 (2006)
3. Hackam, R.: Outdoor HV Composite Insulator. *IEEE Transactions on Dielectrics and Electrical Insulation* 6(5), 557–585 (1999)
4. Giriantari, I.A.D., Blackburn, T.R.: Frequency Characteristics of PD Waveforms on Polluted Composite Insulators Surfaces. In: *Proceedings of 2001 International Symposium on Electrical Insulating Materials*, Himeji, pp. 391–394 (2001)
5. La, O.D.A., Gorur, R.S.: Electrical Performance of Non-ceramic Insulators in Artificial Contamination Test: Role of Resting Time. *IEEE Transactions on Dielectrics and Electrical Insulation* 3(6), 827–835 (1996)
6. Liang, X.D., Wang, S.W.: Artificial Pollution Test and Pollution Performance of Composite Insulators. In: *Proceedings of the 11th International Symposium on High Voltage Engineering*, London, UK, pp. 337–340 (1999)
7. Jang, J.S.R.: ANFIS: Adaptive-Network-Based Fuzzy Inference System. *IEEE transactions on systems, man, and cybernetics* 23(3), 154–157 (1993)

8. Zhang, Z.P., Zhang, Y.C., Li, Y.H.: Monitoring and Localizing of PD in HV Transformers Via an Optical Acoustic Sensor. In: Conference Proceedings of the Seventh International Conference on Electronic Measurement & Instruments, pp. 68–75 (2005)
9. Salam, M.A., Aamer, K., Hamdan, A.: Study the Relationship between the Resistance and ESDD of a Contaminated Insulator a Laboratory Approach. In: International Conference on Properties and Applications of Dielectric Materials, vol. 3, pp. 1032–1034 (2003)
10. Montoya, G., Ramirez, I., Montoya, J.I.: Correlation Among ESDD, NSDD and Leakage Current in Distribution Insulators. IEE Proceedings on Generation Transmission and Distribution 151(3), 334–340 (2004)
11. Liu, P.Y., Li, H.X.: Efficient Learning Algorithms for Three-Layer Regular Feed-forward Fuzzy Neural Networks. IEEE Trans. on Neural Networks 15(3), 545–559 (2004)
12. Arslan, A., Kaya, M.: Determination of Fuzzy Logic Membership Functions Using Genetic Algorithms. J. Fuzzy Sets and Systems 118, 297–306 (2001)

Fault Diagnosis of Analog IC Based on Wavelet Neural Network Ensemble

Lei Zuo, Ligang Hou, Wuchen Wu, Jinhui Wang, and Shuqin Geng

VLSI & System Lab, Beijing University of Technology, Beijing 100022, China

Abstract. A new method of analog IC fault diagnosis is proposed in this paper, which is based on wavelet neural network ensemble (WNNE) technique and Adaboost algorithm. This makes the way of the directory be of use in fault, and enhances the validity of the fault diagnosis. Using wavelet decomposition as a tool for extracting feature, Then, after training the WNNE by faulty feature vectors, the fault diagnosis of a radar scanning circuit is implemented with this new method. The simulation results show that the new method is more effective than the traditional wavelet neural network (WNN) method.

Keywords: Wavelet neural network ensemble, Fault diagnosis, Adaboost.

1 Introduction

Fault diagnosis and testing of electronic systems becomes a crucial and complex task in the past two decades. The fault diagnosis of analogue circuits is more difficult than the digital ones [1], because of their poor fault models, noise, nonlinearity and tolerance effects. Neural networks have been widely used for fault diagnosis of analog circuits because their strong capability in tackling classification and nonlinear problem in recent years [2-4], but the single network has poor generalization ability.

In 1990, Hansen and Salamon proposed neural networks ensemble [5]. The main idea of neural network ensemble is training many neural networks and then combining their outputs. Therefore, this technique has been applied in many fields successfully, such as fault diagnosis, Speaker Recognition and so on [6-7]. However, little attention has been paid to the use of this new technique in analog circuit fault diagnosis.

Therefore, Wavelet Neural Network ensemble (WNNE) was applied to analog circuit fault diagnosis in this paper. It produces a set of Neural Network ensemble classifiers by use of Adaboost algorithm. It can significantly improve the generalization capability than the single network. In order to overcome shortcomings of easily traps to a local optimum, slow velocity of convergence and networks stagnation, a good training algorithm combining PSO with WNNE is proposed in this paper.

The results have been proved that neural network ensembles are feasible and effective to analog circuit fault diagnosis.

2 Implement Neural Network Ensemble

Two of the commonly used techniques for constructing Ensemble classifiers are boosting and bagging [8-9]. As the most popular Boosting method, Adaboost be attributed to its

ability to enlarge the margin, which could enhance the generalization capability. Main steps of the Adaboost algorithm:

1):Input: a set of training samples with labels $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, $x_i \in X, y_i \in Y = \{1, 2, \dots, k\}$, k is the times of iteration. T is the number of cycles. $D(i)$ is the distribution.

2): Initialize: the weights of training samples: $w_i^t = D(i) / k - 1, i = 1, 2, \dots, N_0$,

$$D(i) = 1/N_0 .$$

3): for $t = 1, 2, \dots, T$

$$W_i^t = \sum_{y \neq y_i} w_{i,y}^t, q_t(i, y) = \frac{w_{i,y}^t}{W_i^t}, y \neq y_i, D_t(i) = \frac{W_i^t}{\sum_{i=1}^N W_i^t}$$

4):Train weak learner, classifier $h_t: X \times Y \rightarrow [0, 1]$;

5): Calculate the trade off measure between accuracy and diversity of h_t :

$$\epsilon_t = \frac{1}{2} \sum_{i=1}^N D_t [1 - h_t(x_i, y_i) + \sum_{y \neq y_i} q_t(i, y) h_t(x_i, y)] \tag{1}$$

6): Set the weight of component classifier $h_t: \beta_t = a_t / (1 - a_t)$;

7): Update the weights of training samples: $w_{i,y}^{t+1} = w_{i,y}^t \beta_t^{(1/2)[1+h_t(x_i, y_i) - h_t(x_i, y)]}$

8): Output:

$$h_f(x) = \arg \max_{y \in Y} \sum_{t=1}^T h_t(x, y) \log(1/\beta_t) \tag{2}$$

$h_t(x, y)$ is a classifier, q_t is a weight function. q_t denotes ‘‘Easy’’ samples that are correctly classified ht get lower weights, and ‘‘hard’’ samples that are misclassified get higher weights. Thus, Adaboost focuses on the samples with higher weights, which seem to be harder for Component Learn. This process continues for T cycles, and finally, Adaboost linearly combines all the component classifiers into a single final hypothesis h_f .

3 Analog Circuit Fault Diagnosis Model Based on WNNE

In this paper we use the wavelet transform to extract appropriate feature vectors from the signals sampled from the circuit under test (CUT) under various faulty conditions and optimal feature vectors are selected to train the wavelet neural networks by gradient. And wavelet neural networks ensemble identifies the fault class by the output of the network trained with various fault patterns.

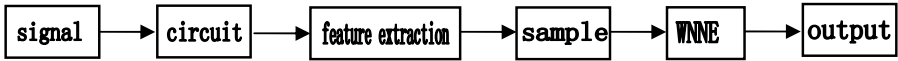


Fig. 1. Fault model of analog circuit

3.1 Wavelet Neural Network

Neural networks have many features suitable for fault diagnosis. WNN is constructed based on wavelet analysis, which has similar structure of feed-forward neural networks. Three-layer WNN is embedded with wavelet functions as hidden layer neurons, which take wavelet space as feature space of pattern recognition. This is a multi-layer feedback architecture with wavelet, allowing the minimum time to converge to its global maximum. The WNN employs a wavelet base rather than a sigmoid function, which discriminates it from general back propagation neural networks.

The function of mapping can be expressed as:

$$f(x) = \sum_{o=1}^h \sum_{j=1}^m \omega_o \frac{1}{\sqrt{|a_o|}} \psi_{j_o} \left(\frac{\sum_{i=1}^n x_{ij} - b_o}{a_o} \right) \tag{3}$$

$\omega_o (o=1,2,\dots,h)$ is output of hidden layer neurons; ψ_{j_o} is the wavelet bases.

Networks have three parameters to be trained: Output weight ω , translation factors a and dilation factors b .

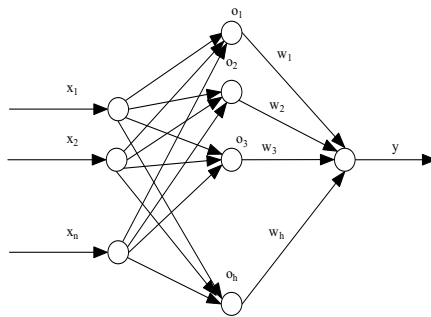


Fig. 2. Framework of wavelet neural network

3.2 Particle Swarm Optimization

Particle swarm optimization (PSO) is a population based stochastic optimization technique developed by Eberhart and Kennedy in 1995 [10]. The PSO algorithm is inspired by social behavior of bird flocking or fish schooling; it has been applied successfully to function optimize, game learning, data clustering, image analysis, and NN training. The PSO is similar to the concept of the mutation operator found in conventional genetic algorithm, but differs from the typical mutation operator in that it

is not entirely random. The studies showed that the PSO has more chance to “fly” into the better solution areas more quickly, so it can discover reasonable quality solution much faster than other evolutionary algorithms. But it did not possess the ability to perform a fine search to improve upon the quality of the solution as the number of generations was increased. PSO is initialized with a group of random particles (solutions) and then searches for optima by updating generations. In the every iteration, each particle is updated by following two "best" values. The first one is the position vector of the best solution (fitness) this particle has achieved so far. The fitness value is also stored. This position is called *pbest*. Another "best" position that is tracked by the particle swarm optimizer is the best position, obtained so far, by any particle in the population. This best position is the current global best and is called *gbest*. After finding the two best values, the particle updates its velocity and position according to equations (4) and (5).

$$v_{iD}^{k+1} = w \times v_{iD}^k + c_1 \times rand(\cdot) \times (p_{iD} - x_{iD}^k) + c_2 \times rand(\cdot) \times (p_{gD} - x_{iD}^k) \quad (4)$$

$$x_{iD}^{k+1} = x_{iD}^k + v_{iD}^{k+1} \quad (5)$$

3.3 Proposed Algorithm

The fault diagnosis method based on wavelet neural network ensemble identifies the fault class by the output of the wavelet neural network ensemble trained with various fault patterns. The WNN ensemble is built in the level of multi-class classifiers and the structure show in Fig.3. We obtain the final decision from the decision results of many multi-class classifiers via an appropriate aggregating strategy of the WNN ensemble. We propose a novel Adaboost-WNN ensemble algorithm in following section.

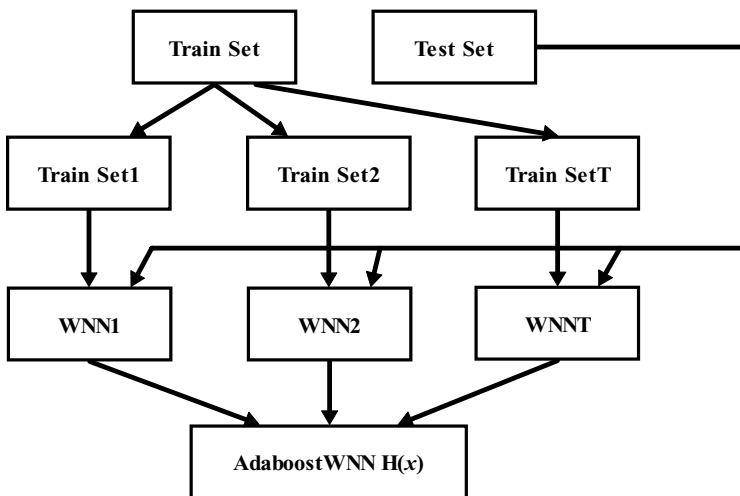


Fig. 3. The structure of WNN ensemble

In this paper, a set of training samples with labels $D = \{(x_i, y_i) | i = 1, 2, \dots, N\}$, morlet wavelet is used as stimulation function of hidden layer: $\psi(x) = \cos(1.75x)e^{-x^2/2}$.

The error performance function is given by:

$$J = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^m (y_{j,i}^d - y_{j,i})^2 \tag{6}$$

where N is the total number of training patterns, and $y_{j,i}^d$ and y_{ij} are the desired and real outputs respectively.

$h_i(x, y)$ is a classifier:

$$h_i(x_i, y_i) = f_{y_i}(x_i), \quad h_i(x_i, y) = f_y(x_i), \quad y \neq y_i$$

The training process of WNNE is performed as following:

1): Create initial population of individuals according to the initiation strategy. Output weight ω translation factors a and dilation factors b are in(0,1). r_1 and r_2 are random numbers between 0 and 1. c_1 is the self confidence (cognitive) factor; c_2 is the swarm confidence (social) factor. Usually c_1 and c_2 are in the range from 1.5 to 2.5; w is the inertia factor that takes values downward from 0.7 to 0.4 according to the iteration number.

2): Calculate the fitness function by equation (6).

3): To minimize the fitness function in (4), the weights and coefficients a and b can be updated using the equations (4) and (5).

4): Repeat step 2) to step 3) until some constraint condition is satisfied, then stop and the desired individuals are obtained.

5): Training by Adaboost, $T=30$, record the classifier $h_i(x, y)$ every cycles, if some constraint condition is satisfied, then stop and output by equation (2),the outputs of the WNNE will show the fault patterns.

4 Example Circuits and Faults

The circuit in Fig. 4 is a scanning circuit of radar [11]. The scanning circuit is consist of monostable circuit 1 and sawtooth wave circuit 2.The faults associated with this circuit are assumed to Z_1 turn off, Z_2 turn off D_1 , The fault classes are binary encoded that digit 0 indicates fault free and 1 faulty for corresponding components. (Shown in Tab.1)

Wavelet transform is a mathematical operation that decomposes input signal simultaneously into time and frequency components. Wavelet analysis can extract signature of signals and compress the signature data at the same time, which is a powerful tool of non-stationary signal processing. The circuit simulation software is Pspice. Wavelet transform is use to extract signal feature of OUT2, OUT3, OUT4 and OUT4, OUT5, OUT6, OUT7, OUT8. The usually used orthogonal wavelets are Harr wavelet and Db3 wavelet.

The Db3 wavelet packet transform is applied to decompose all the samples, there are 4 wavelet parameters $\{d_1, d_2, d_3, c_3\}$, C_3 represent the aggregation of c_3 ,

$D_j(j=1,2,3)$ represent the aggregation of d_j , The network has 4 inputs $\{D_1, D_2, D_3, C_3\}$. In the paper, the signal feature of circuit 1 is 12, and circuit 2 has 16.

The work must have train 2 networks for circuit 1 and circuit 2. The neural network 1 has three layers: input layer, hidden layer and output layer. Input layer has 12 neurons, hidden layer has 30 neurons. Output layer has 5 neurons. The neural network 2 has three layers: Input layer has 16 neurons, hidden layer has 30 neurons. Output layer has 7 neurons.

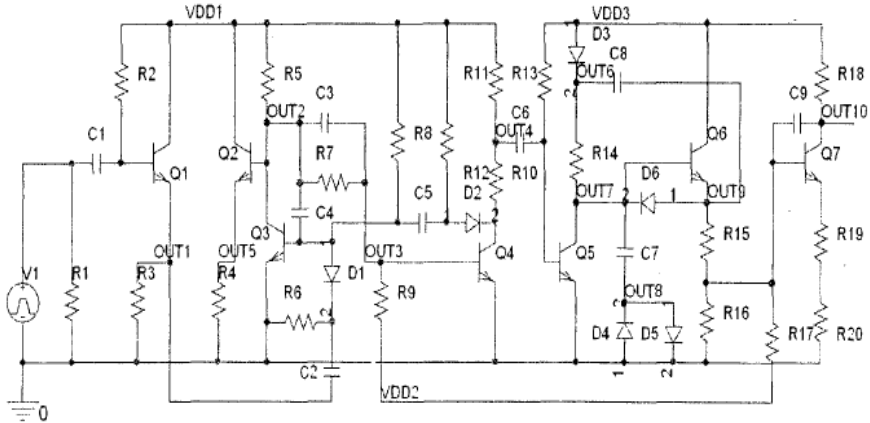


Fig. 4. Scanning circuit of radar

Table 1. Fault mode

mode	encoded
Transistor Q_2 open	0 0 0 0 0 0 0 0 0 0 0 1
Transistor Q_2 breakdown	0 0 0 0 0 0 0 0 0 0 1 0
Transistor Q_3 open	0 0 0 0 0 0 0 0 0 1 0 0
Detector D_1 breakdown	0 0 0 0 0 0 0 0 1 0 0 0
Transistor Q_3 breakdown	0 0 0 0 0 0 0 1 0 0 0 0
Transistor Q_4 breakdown	0 0 0 0 0 0 1 0 0 0 0 0
Transistor Q_4 breakdown	0 0 0 0 0 1 0 0 0 0 0 0
Transistor Q_5 open	0 0 0 0 1 0 0 0 0 0 0 0
Transistor Q_5 breakdown	0 0 0 1 0 0 0 0 0 0 0 0
Detector D_3 breakdown	0 0 1 0 0 0 0 0 0 0 0 0
Detector D_4 or D_5 breakdown	0 1 0 0 0 0 0 0 0 0 0 0
Detector D_6 breakdown	1 0 0 0 0 0 0 0 0 0 0 0

We have 100 groups of data for each fault mode, totally 500 data for circuit 1 by Monte-Carlo. 100 groups of data for each fault mode, totally 700 data for circuit 2 by Monte-Carlo. And we also have 300 and 350 test data for circuit 1 and circuit 2 by using the same method.

In order to compare diagnosis ability of wavelet neural network ensemble, this paper also trained wavelet neural network, from Table 2 we can know that the average classification accuracy rate of the two methods.

We can know that the average classification accuracy rate of the multi-classifiers based on WNNE is 98% which is higher obviously than WNN is able to properly classify 93.1% of the test patterns. for the Proposed method using Adaboost creates a collection of diversity component classifiers which improve Classifier's classification and generalization ability by maintaining a set of weights over training samples and Adaptively adjusting these weights after each adaboosting iteration: the weights of the training samples which are misclassified by current component classifier will be increased while the weights of the training samples which are correctly classified will be decreased. Classification results for the experiment prove the performance improvement of the proposed WNNE fault diagnosis method is better than the WNN method.

Table 2. The result of WNN and WNNE

mode	Accuracy of testing samples by WNN	Accuracy of testing samples by WNNE
Transistor Q ₂ open	92.7%	97.3%
Transistor Q ₂ breakdown	91.7%	96.7%
Transistor Q ₃ open	90.7%	97.3%
Detector D ₁ breakdown	91.7%	97.3%
Transistor Q ₃ breakdown	92.0%	98.0%
Transistor Q ₄ breakdown	92.7%	98.3%
Transistor Q ₄ breakdown	93.3%	98.3%
Transistor Q ₅ open	94.0%	98.3%
Transistor Q ₅ breakdown	94.0%	98.7%
Detector D ₃ breakdown	94.3%	98.7%
Detector D ₄ or D ₅ breakdown	95.0%	98.7%
Detector D ₆ breakdown	95.0%	98.7%

5 Conclusion

We have presented our first results in the development of a new technique for fault diagnosis of analogue circuits. The method we propose a novel algorithm combining wavelet analysis, WNN and Adaboost. This paper used wavelet analysis to extract the signal feature. The optimal feature sets are then used to train the wavelet neural network, and then using Adaboost creates wavelet neural network ensemble. In applying WNNE to the diagnosis of scanning circuit of radar circuits, as described, we have demonstrated the implementation of the method and a set of results. These results vindicate the technique is more effective than WNN technique.

References

1. Aminian, M., Aminian, F.: Neural-network based analog-circuit fault diagnosis using wavelet transform as preprocessor. *IEEE Transaction on Circuits and Systems* 47, 151–156 (2000)
2. Martin, H.T.: Back propagation neural network design. China Machinery Press, Beijing (2002)

3. Stopjakova, V., Micusik, D., L'Benuskova, et al.: Neural networks-based parametric testing of analog IC. In: Proc. of IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems, pp. 408–416 (2002)
4. He, Y., Ding, Y., Sun, Y.: Fault diagnosis of analog circuits with tolerances using artificial neural networks. In: Proc. IEEE APCCAS, Tianjin, pp. 292–295 (2000)
5. Hansen, L.K., Peter, S.: Neural network ensemble. *IEEE Trans Pattern Analysis and Machine Intelligence* 12, 993–1001 (1990)
6. James, W., Taylor, Roberto, B.: Neural network load forecasting with weather ensemble predictions. *IEEE Transaction on Power Systems* 17, 626–632 (2002)
7. Abdullah, M.H.L.B., Ganapathy, V.: Neural network ensemble for financial trend prediction. In: Proceedings of TENCON, vol. 3, pp. 157–161 (2000)
8. Leo, B.: Bagging predictors. *Machine Learning* 24, 123–140 (2000)
9. Yoav, F., Robert, E.S.: Experiments with a new boosting algorithm. In: Proceedings of International Conference on Machine Learning, pp. 148–156. Morgan Kaufmann, San Francisco (1996)
10. Kennedy, J., Eberhart, R.C.: Particle swarm optimization. In: Proc. IEEE Int. Conf. Neural Network IV, Perth, Australia, pp. 1942–1948 (1995)
11. Luo, Z.Y.: Research on intelligent fault diagnosis techniques for radar system. Xi'an College of Automation Northwestern Polytechnical University, pp. 75–80 (2006)

Dynamic Neural Network-Based Fault Detection and Isolation for Thrusters in Formation Flying of Satellites

Arturo Valdes¹, K. Khorasani¹, and Liying Ma²

¹ Department of Electrical and Computer Engineering,
Concordia University, Montreal, QC H3G 1M8, Canada

² Department of Applied Computer Science, Tokyo Polytechnic University,
1583, Iiyama, Atsugi, Kanagawa, Japan 243-0297
kash@ece.concordia.ca, maly@cs.t-kougei.ac.jp

Abstract. The objective of this paper is to develop a dynamic neural network-based fault detection and isolation (FDI) scheme for satellites that are tasked to perform a formation flying mission. Specifically, the proposed FDI scheme is developed for Pulsed Plasma Thrusters (PPT) that are considered to be used in satellite's Attitude Control Subsystem (ACS). By using the relative attitudes of satellites in the formation our proposed "High Level" fault diagnoser scheme can detect a pair of thrusters that are faulty. This high level diagnoser however cannot isolate the faulty satellite in the formation. Towards this end, a novel "Integrated" dynamic neural network (DNN)-based FDI scheme is proposed to achieve both fault detection and fault isolation of the formation flying of satellites. This methodology involves an "optimal" fusion of the "High Level" FDI scheme with a DNN-based "Low Level" FDI scheme that was recently developed by the authors. To demonstrate the FDI capabilities of our proposed schemes various fault scenarios are simulated and a comparative study among the techniques is performed.

Keywords: Dynamic neural networks, Fault detection and isolation, Attitude control subsystem, Pulsed plasma thrusters, Formation flying of satellites.

1 Introduction

Development of a fault detection and isolation (FDI) scheme for unmanned space vehicles is a challenging problem. Traditionally, spacecraft sends periodic batch of data to ground stations where the data is analyzed in order to determine the health status of various subsystems. When a fault is detected, additional analyses must be performed to isolate the fault. This process is a time-consuming task which is also very costly. Due to these and other considerations, there is a real interest in developing autonomous fault diagnostic approaches for on-board spacecraft subsystems especially for the attitude control subsystem (ACS) of formation flying of multiple satellites. Literature on FDI of spacecraft provides various studies on faulty components of the ACS of single spacecraft ([1]-[7]). However, for all practical purposes there is no work on FDI of formation flying spacecraft missions.

Malfunctions in any of the ACS components can affect the performance of the mission. Therefore, early detection of faults and isolation of faulty components become extremely important. To the best of authors' knowledge, development of an FDI module for detecting and isolating faults in Pulsed Plasma Thruster (PPT) in formation flying spacecraft has not been investigated in the literature.

In this paper, an FDI scheme based on dynamic neural networks (DNN) is proposed and developed. Based on relative attitudes of the formation flying spacecraft, our proposed FDI scheme is capable of detecting the spacecraft that is affected by a fault. An integrated FDI scheme is proposed that is composed of a "High Level" FDI scheme and a "Low Level" FDI scheme that has recently been developed in [8]. The resulting "Integrated" FDI scheme does take advantage of the strengths of each scheme and at same time minimizes their weaknesses. In order to demonstrate the capabilities of our proposed FDI schemes, a three spacecraft formation flying under various fault scenarios are investigated and analyzed. The results demonstrate that the "Integrated" FDI scheme exhibits improved fault detection and isolation capabilities than either the "Low" or the "High" level FDI schemes individually.

The remainder of the paper is organized as follows: In Section 2, the dynamic neural network-based fault detection and isolation scheme which uses information from the attitude control subsystem of the formation flying spacecraft is developed. In order to evaluate our proposed FDI schemes, simulated fault scenarios are investigated and the results are presented and discussed. In Section 3, the DNN-based FDI scheme developed in [8] is briefly reviewed and its strengths and weaknesses are discussed. The motivations for proposing our integrated FDI scheme are then established. Furthermore, the development of our "Integrated" FDI scheme is presented and its performance is evaluated and compared with the scheme proposed in [12]. In Section 4, conclusions and contributions of our proposed FDI schemes are provided.

2 Formation Flying Fault Detection and Isolation Methodology

In this section an FDI approach for the formation flying mission that is composed of three spacecraft with a leader/follower control architecture is developed. The satellites use the so-called six-independent pulsed plasma thruster (PPT) configuration. In this configuration each PPT only generates a torque about a single axis of the spacecraft where independent control actuation is achieved.

Dynamic neural networks (DNN) are employed to model the relative attitude of followers spacecraft with respect to the leader spacecraft in a formation flying mission. Using this neural network model, residual signals are generated for detecting the existence of faults in the actuators of the followers. An important advantage of this FDI scheme is that only data from the follower's ACS is used to detect abnormalities in the actuators.

Pulsed plasma thrusters (PPTs) are accurate, inexpensive and simple actuators that can be used for different purposes such as station-keeping, attitude control and orbit insertion and drag make-up [9-12]. As shown in Fig. 1, the main components of the PPT are the capacitor, the electrodes, the igniter and the spring. Once the igniter is discharged, the capacitor voltage that appears across the electrodes creates a current which ablates and ionizes the fuel bar into a plasma slug. Finally, the plasma is accelerated by the Lorentz force ($\mathbf{J} \times \mathbf{B}$) due to the discharge current and the magnetic field.

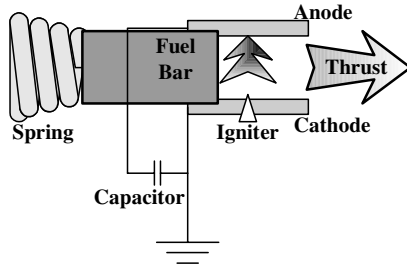


Fig. 1. The schematic of a pulsed plasma thruster (PPT)

The PPT is an electromechanical system where the acceleration process and circuit components can be modeled by the following dynamical system

$$\begin{aligned} \dot{x}_1 &= x_3(t) \\ \dot{x}_2 &= x_4(t) \\ \dot{x}_3 &= \frac{1}{2} \frac{L'_{pe}}{m_0} [x_3(t)]^2 \\ \dot{x}_4 &= \frac{-\left(\frac{1}{C}\right)x_2(t) - \mu_0 \frac{h}{w} x_3(t)x_4(t) - R_T x_4(t) + v(t)}{L_T(t)} \end{aligned} \tag{1}$$

$$T(t) = m_0 f x_3(t)$$

$$c(t) = x_4(t)$$

where $x_1(t)$ is the position, $x_2(t)$ is the capacitor charge, $x_3(t)$ is the velocity, $x_4(t)$ is the current, $v(t)$ is the capacitor voltage, $T(t)$ is the thrust, $c(t)$ is the discharge current, $R_T = R_c + R_e + R_{pe} + R_p$ and $L_T(t) = L_c + L_e + L_{pe}(t)$. Equations (2) and (3) provide expressions for R_p and $L_{pe}(t)$, respectively, and the parameters appearing in (1)-(3) are specified in Table 1. We furthermore assume that $x_1(0) = x_2(0) = x_3(0) = x_4(0) = 0$.

$$R_p = 2.57 \frac{h}{T_e^{\frac{3}{4}} w} \sqrt{\frac{\mu_0 \ln \left[1.24 \times 10^7 \left(\frac{T_e^3}{n_e} \right)^{\frac{1}{2}} \right]}{\tau}} \tag{2}$$

$$L_{pe}(t) = \mu_0 \frac{h}{w} x_1(t) \tag{3}$$

Table 1. Parameters of the Parallel-Plates Ablative PPT Electromechanical Model

Parameter	Description
C	Capacitance (F)
f	Pulse Frequency (Hz)
h	Distance between electrodes (m)
L_c	Internal Inductance of the capacitor (H)
L_e	Inductance due to wires and leads (H)
L_{pe}	Inductance due to current sheet moving down (H)
L'_{pe}	Inductance per unit channel length (Hm^{-1})
L_T	Total circuit inductance (H)
m_0	Mass of plasma at $t = 0$ (kg)
n_e	Electron density (m^{-3})
R_c	Capacitor resistance (Ω)
R_e	Wire and lead resistance (Ω)
R_p	Plasma resistance (Ω)
R_{pe}	Electrode resistance (Ω)
R_T	Total circuit resistance (Ω)
T_e	Electron temperature
w	Width of electrodes (m)
μ_0	Magnetic permeability of free space ($\text{WbA}^{-1}\text{m}^{-1}$)
τ	Characteristic pulse time (s)

During normal/healthy operations, only the electrical variables and temperature of the PPT thrusters are measurable. As indicated in [8], typically PPT thrusters are grouped into pairs sharing the same capacitor and therefore both PPTs cannot generate thrust pulses at the same time. To obtain a full three-axis control, a minimum of four thrusters is needed. However, in this paper, we consider the so-called six-independent PPT configuration (refer to Fig. 2 for details). In the above configuration, each thruster only generates a torque about a single axis of the spacecraft where independent control actuation is achieved. By using this configuration the applied torques in the $+x, -x, +y, -y, +z$ and $-z$ directions are performed by the thrusters $PPT_1, PPT_2, PPT_3, PPT_4, PPT_5$ and PPT_6 , respectively.

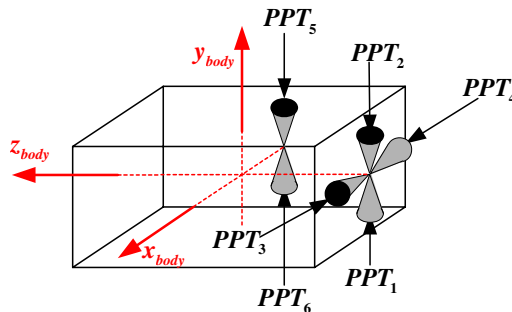


Fig. 2. The six-independent PPT configuration

By means of sensors the leader spacecraft (S/C_l) can measure its absolute angular rotations and velocities. For the follower spacecraft (i.e. S/C_{f_1} and S/C_{f_2}) it is necessary to measure the relative attitudes with respect to the leader. Beside these measurements, the number of pulses that are generated by each PPT and the instant (time) when pulses are generated are also recorded by each spacecraft. Table 2 shows the set of variables that are defined above (where l represents the leader spacecraft and f,j represent the j -th follower spacecraft with $j=1$ or 2).

Table 2. Attitude Variables and Sequence of Pulses of the j -th Follower Spacecraft

Variable	Description
${}^{f,j}q_1$: angular rotation about the x -axis (S/C_{f_j} w.r.t. S/C_l)
${}^{f,j}q_2$: angular rotation about the y -axis (S/C_{f_j} w.r.t. S/C_l)
${}^{f,j}q_3$: angular rotation about the z -axis (S/C_{f_j} w.r.t. S/C_l)
${}^{f,j}\Delta\omega_x$: angular velocity about the x -axis (S/C_{f_j} w.r.t. S/C_l)
${}^{f,j}\Delta\omega_y$: angular velocity about the y -axis (S/C_{f_j} w.r.t. S/C_l)
${}^{f,j}\Delta\omega_z$: angular velocity about the z -axis (S/C_{f_j} w.r.t. S/C_l)
${}^{f,j}T_{PPT1/PPT2}$: sequence of pulses about the x -axis (S/C_{f_j} w.r.t. S/C_l)
${}^{f,j}T_{PPT3/PPT4}$: sequence of pulses about the y -axis (S/C_{f_j} w.r.t. S/C_l)
${}^{f,j}T_{PPT5/PPT6}$: sequence of pulses about the z -axis (S/C_{f_j} w.r.t. S/C_l)

2.1 Design of Neural Network FDI Scheme

The neural networks considered in this paper is a multilayer perceptron network with dynamics neurons. As presented in [13]-[18] these special neurons allow the network to achieve dynamics properties. Fig. 3 shows the general structure of the so-called Dynamic Neuron Model (DNM) [15]-[26].

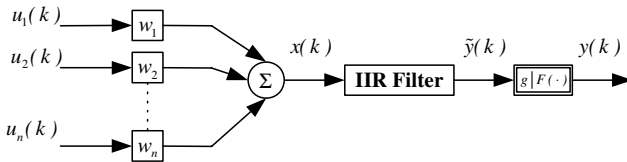


Fig. 3. A dynamic neuron model

The set $[u_1(k), u_2(k), \dots, u_n(k)]^T$ and $W = [w_1, w_2, \dots, w_n]^T$ are the input and weight vectors, respectively. An Infinite Impulse Response (IIR) Filter is introduced to generate dynamics in the neuron such that the activation of a neuron depends on its internal states [15]-[18]. The block $g|F(\cdot)$ is the activation function of the neuron. The

parameter g , is the slope of the nonlinear activation function represented by $F(\cdot)$. The dynamic model of the above neuron is described by the following set of equations:

$$\begin{aligned}
 x(k) &= \sum_{i=1}^n w_i u_i(k) \\
 \tilde{y}(k) &= -\sum_{i=1}^r a_i \tilde{y}(k-i) + \sum_{i=0}^r b_i x(k-i) \\
 y(k) &= F(g \cdot \tilde{y}(k))
 \end{aligned}
 \tag{4}$$

where the signal $x(k)$ represents the input to the filter, the coefficients $a_i, i = 1, 2, \dots, r$ and $b_i, i = 0, 1, \dots, r$ are the feedback and feed-forward filter parameters, respectively, and r is the order of the filter. Finally, $\tilde{y}(k)$ represents the output of the filter which is the input to the activation function.

2.2 Training and Generalization Architectures for the ‘‘High Level’’ FDI Scheme

In order to collect data for the training phase, different fault-free formation flying missions are simulated. Fig. 4 shows the schematic representation of the proposed DNN that is used for the x -axis (i.e. roll angle) during the training phase. From this figure one can see that the sequence of pulses about the x -axis generated by the pair of thrusters PPT_1/PPT_2 and the angular rotations about the three axes are presented to the DNN_{roll} . The output of the network is the estimated angular velocity about the

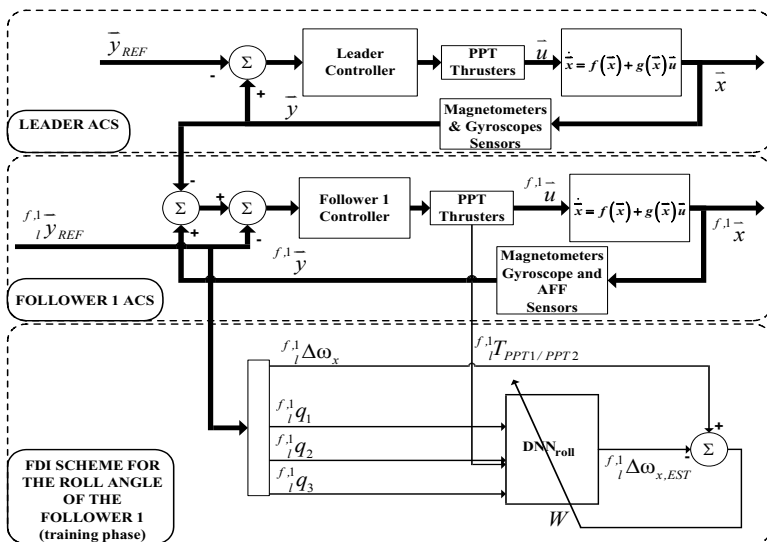


Fig. 4. Identification model during the training phase

x -axis. By comparing the output of the DNN_{roll} with the measured angular velocity about the x -axis, the estimation error is calculated and back-propagated through various layers updating the network parameters W . The training algorithm used here is the Extended Dynamic Back Propagation (EDBP) algorithm [18]. The DNN_{roll} is trained until a termination criterion ($t.c.$) is fulfilled. Here the criterion used is the mean square error (mse).

The training process above is also used for the other two DNNs (i.e. DNN_{pitch} and DNN_{yaw}). Each DNN has a 4-10-1 structure (four neurons in the input layer, ten neurons in the hidden layer and one neuron in the output layer) with second order Infinite Impulse Response (IIR) filters and hyperbolic tangent sigmoid and linear activation functions for the neurons in the hidden and output layers, respectively.

Once the training phase is completed, the parameters of the dynamic neural networks are fixed and the validation phase is initiated. Data generated from missions that are different from those used for the training purpose is presented to the DNN. By comparing the estimated angular velocity of the DNN with the measured angular velocity the representation capabilities of the network are analyzed. Fig. 5 shows the DNN architecture that is used for the roll angle during the validation phase.

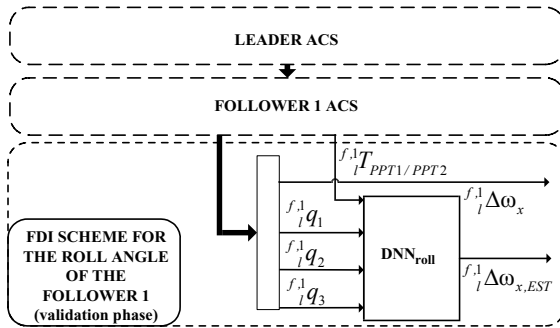


Fig. 5. Identification model during the validation phase

The next phase deals with the calculation of a threshold function and an FDI evaluation criterion. This threshold will be used for determining the health status of the pair of PPT thrusters. The value of the threshold is calculated by using the healthy data collected from the simulated formation flying missions. The calculation of the $Threshold_{roll}$ is performed by using the mathematical expression that is given below

$$Threshold_{roll} = \frac{\sum_{l=1}^6 SAE_{roll}(l)}{6} + \sigma_{roll} \left(\max(SAE_{roll}(l))_{l=1}^6 - \frac{\sum_{l=1}^6 SAE_{roll}(l)}{6} \right) \quad (5)$$

where $SAE_{roll}(l)$ is the Sum Absolute Error of the data set $l = 1, 2, \dots, 6$ generated from the six different missions. The coefficient σ is a constant which is used to adjust the

sensitivity of our FDI scheme. Equation (4) is also used for calculating the threshold values for the pitch and yaw angles.

The simulated missions require that the S/C_{f1} and S/C_{f2} spacecraft rotate from an initial angular position (i.e. $[0^\circ, 0^\circ, 0^\circ]$) until they reach the desirable attitudes (i.e. the reference attitudes). After calculating the SAE values and using the coefficients $\sigma_{roll} = 1.208$, $\sigma_{pitch} = 2.450$, and $\sigma_{yaw} = 1.032$, the thresholds obtained are: $Threshold_{roll} = 135.00$, $threshold_{pitch} = 50.00$, and $Threshold_{yaw} = 76.00$, respectively.

Our proposed FDI scheme for detecting a single axis can be represented as shown in Fig. 6. In this scheme, the attitudes of the S/C_{f1} are applied to the DNNs and the estimated angular velocities are compared with the actual measurements and the corresponding SAE values are calculated. Finally, the SAE values are compared with the corresponding thresholds and the health status of the three pairs of thrusters are determined.

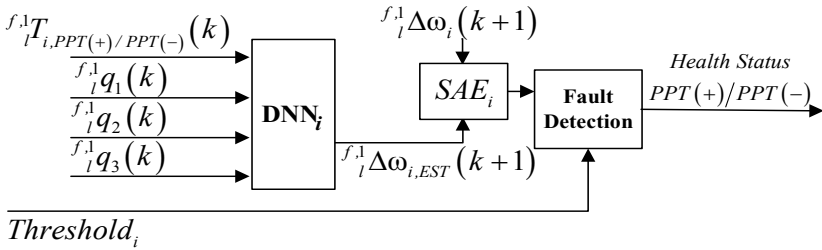


Fig. 6. FDI scheme for the S/C_1 satellite in the formation (subscript i denotes the i -th angle (roll, pitch or yaw), $PPT_i(+)$ denotes the PPT thruster that generates the thrust in the positive direction of the i -th axis (i.e. PPT_1, PPT_3 and PPT_5) and ${}_{5}PPT_i(-)$ denotes the PPT thruster that generates the thrust in the negative direction of the i -th axis (i.e. PPT_2, PPT and PPT_6).

2.2 High Level FDI Scheme Simulations Results

In this section simulations are conducted for evaluating our proposed DNN-based High Level FDI scheme for formation flying missions. The fault types considered are as following:

- **Fault Type 1:** Loss of elasticity is a spring’s failure which affects the deflection of the spring reducing the pressure applied to the propellant bar. This type of failure may change the amount of propellant mass consumed in each pulse.
- **Fault Type 2:** The ablation process transforms the solid propellant into the exhaust plasma, but small portions of the propellant may not be transformed, resulting in particles which are added to the inner face of the electrodes. After several pulses, this situation may lead to degradations of the PPT performance.

- **Fault Type 3:** Due to wear and tear, conductivity of the wires, capacitor and electrodes may decrease. As a consequence of this, the amount of thrust produced may be changed in an unpredictable manner.

To evaluate the performance of our proposed FDI scheme three formation flying missions that are affected by the above faults (i.e. faults affecting the PPTs of the S/C_{f1}) are simulated. For these cases, the reference attitudes are as follows: mission (a) ($[25^\circ, 30^\circ, 40^\circ]$), mission (b) ($[20^\circ, 35^\circ, 45^\circ]$), and mission (c) ($[25^\circ, 35^\circ, 45^\circ]$). Table 3 shows the type of faults that are injected and the PPTs that are affected by the faults and their severity.

Table 3. General Description of the Simulated Faulty Cases

mission	Fault type	Faulty PPT	Severity
(a)	Type 1	PPT_2 of S/C_{f1}	The thrust generated by PPT_2 is decreased by 15%
(b)	Type 2	PPT_3 of S/C_{f1}	The thrust generated by PPT_3 is increased by 15%
(c)	Type 3	PPT_6 of S/C_{f1}	The thrust generated by PPT_6 is decreased by 15%

The SAE values and the Health Status that are obtained for the follower S/C_{f1} are presented in Table 4.

Table 4. Health Status Results

mission	SAE_{roll}	SAE_{pitch}	SAE_{yaw}	Health Status
(a)	382.12	41.53	45.16	PPT_1/PPT_2 is detected as the faulty pair
(b)	100.21	57.88	39.75	PPT_3/PPT_4 is detected as the faulty pair
(c)	132.84	38.40	279.98	PPT_3/PPT_6 is detected as the faulty pair
<i>Threshold:</i>	135.00	50.00	76.00	

According to our simulation results, low severity faults are not observable in the attitudes of the spacecraft because the ACS can fulfill the mission requirements by changing the sequence and the number of pulses generated by the PPTs. Table 5 presents the number of pulses that are generated by S/C_{f1} and S/C_{f2} during the mission (b).

Table 5. Amount of Pulses Generated by the Followers S/C_{f1} and S/C_{f2} During the Mission (b)

Spacecraft	PPT_1/PPT_2 pulses	PPT_3/PPT_4 pulses	PPT_3/PPT_6 pulses
S/C_{f1}	78/510	232/238	228/248
S/C_{f2}	19/21	88/74	68/78

The six PPTs of the S/C_{f1} generated more pulses than the PPTs of the S/C_{f2} to perform the same rotational maneuver. Due to the fact that the operational lifetime of the PPT thrusters is determined by the amount of generated pulses (i.e. number of capacitor’s discharges), this unplanned extra generation of pulses can reduce the lifetime of the formation flying mission and should be avoided and eliminated.

3 Integrated Fault Detection and Isolation Scheme

In the previous section, we developed a “High Level” FDI scheme that by using the relative attitude variables abnormal spacecraft’s behavior can be detected and the pair of thrusters where the fault is injected can be identified. Unfortunately, the “High Level” FDI scheme cannot isolate the faulty actuator.

On the other hand, by utilizing a “Low Level” DNN-based FDI scheme for the PPT thrusters (as proposed in [8]) we can analyze the health status of the six thrusters pulse by pulse. Experimental results demonstrate that for these three types of faults, especially for the fault type 2 the utilization of a single fixed threshold value affects the reliability of our FDI approach.

The integrated FDI scheme uses the “High Level” approach for detecting which pair of thrusters is healthy and which one is faulty. Once the faulty thruster pair is identified, and based on the cause-effect relationships, one can identify the possible effect of the fault on both PPTs. With this information, different threshold values (i.e. “lower threshold” and “upper threshold”) can be determined. By applying these thresholds to the “Low Level” approach one can determine which thruster is faulty, and more specifically, which one of the generated pulses is faulty. Fig. 7 shows the schematic representation of our “Integrated” FDI scheme. The “High Level”

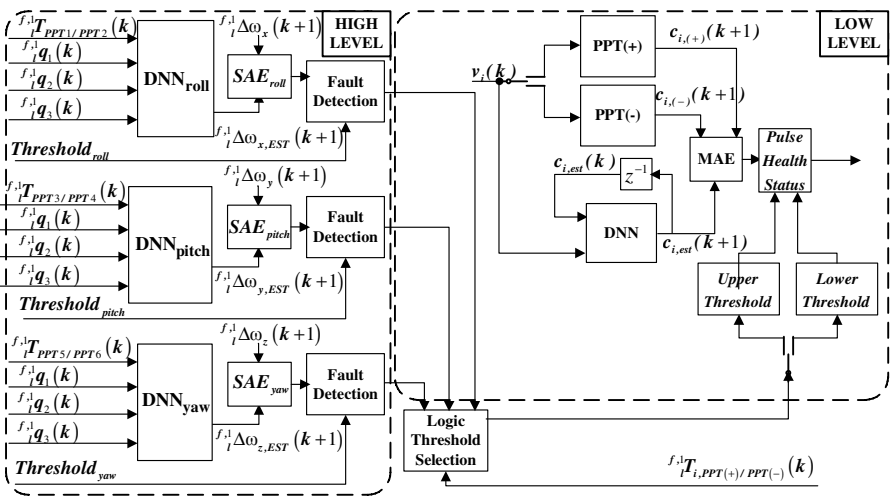


Fig. 7. Our proposed ‘Integrated’ FDI scheme for the S/C_i satellite in the formation flying (the subscript i in the “Low Level” section of the scheme represents the i -th axis (i.e. roll, pitch or yaw angles)).

FDI scheme detects the faulty pair of thrusters and the Logic Threshold Selection block counts the number of pulses that are generated by each PPT of the faulty pair and determines which threshold must be applied to each PPT. Finally, the “Low Level” FDI scheme analyzes pulse by pulse the health status of both PPTs and detects the faulty pulses that are generated by the PPTs.

In order to demonstrate the performance of our proposed “Integrated” FDI scheme six formation flying missions are simulated. The specifications of these missions are presented in Table 6.

Table 6. Specifications of the Simulated Faulty Missions for Evaluating the Performance of Our Proposed “Integrated” FDI Scheme

mission	reference	Faulty type (severity)	Faulty PPT	Occurrence time
1	[50°, 30°, 40°]	Type 2 (incremental)	PPT_1 of S/C_{fl}	t = 10 sec
2	[25°, 30°, 40°]	Type 2 (incremental)	PPT_3 of S/C_{fl}	t = 0 sec
3	[20°, 30°, 40°]	Type 1 (incremental)	PPT_5 of S/C_{fl}	t = 350 sec
4	[20°, 35°, 45°]	Type 1 (incremental)	PPT_6 of S/C_{fl}	t = 850 sec
5	[55°, 30°, 45°]	Type 3 (incremental)	PPT_4 of S/C_{fl}	t = 500 sec
6	[45°, 20°, 35°]	Type 3 (incremental)	PPT_2 of S/C_{fl}	t = 600 sec

Table 7 shows the “High Level” FDI results for the six missions. According to these results, we have positively detected the faulty thruster in all the simulated missions.

Table 7. Results of the “High Level” FDI Scheme for the Six Simulated Faulty Missions

mission	SAE_{roll}	SAE_{pitch}	SAE_{yaw}	Health Status
1	264.96	46.28	59.32	PPT_1/PPT_2 of S/C_{fl} is detected as the faulty pair
2	90.35	53.28	33.45	PPT_3/PPT_4 of S/C_{fl} is detected as the faulty pair
3	118.26	33.08	135.49	PPT_5/PPT_6 of S/C_{fl} is detected as the faulty pair
4	88.23	48.58	172.86	PPT_5/PPT_6 of S/C_{fl} is detected as the faulty pair
5	119.71	53.65	46.26	PPT_3/PPT_4 of S/C_{fl} is detected as the faulty pair
6	196.86	29.05	60.64	PPT_1/PPT_2 of S/C_{fl} is detected as the faulty pair
<i>Threshold:</i>	135.00	50.00	76.00	

Based on the results shown in [8] and simulations performed in this paper we determined that the optimal values for the lower and upper Thresholds are 0.0300 and 0.0370, respectively. Table 8 presents the results given by the Logic Threshold Selection block for the six simulated formation flying missions.

Table 8. Threshold determination for the six simulated faulty missions

mis- sion	$PPT_i(+)$ (Number of Pulses; Threshold value)	$PPT_i(-)$ (Number of Pulses; Threshold value)
1	PPT_1 (138; Lower Threshold)	PPT_2 (885; Upper Threshold)
2	PPT_3 (126; Lower Threshold)	PPT_4 (577; Upper Threshold)
3	PPT_5 (1454; Upper Threshold)	PPT_6 (46; Lower Threshold)
4	PPT_5 (50; Lower Threshold)	PPT_6 (1892; Upper Threshold)
5	PPT_3 (202; Lower Threshold)	PPT_4 (436; Upper Threshold)
6	PPT_1 (143; Lower Threshold)	PPT_2 (382; Upper Threshold)

The “Low Level” FDI scheme uses the threshold values that are determined by the Logic Selection Threshold block to detect the faulty pulses of the faulty actuator. Table 9 shows the results obtained.

Table 9. Results of the “Low Level” FDI Scheme for the Six Simulated Faulty Missions

mission	PPT	actual/detected healthy pulses	actual/detected faulty pulses
1	PPT_1 :	0/0	138/138
	PPT_2 :	885/885	0/0
2	PPT_3 :	0/0	126/126
	PPT_4 :	577/577	0/0
3	PPT_5 :	32/32	1422/1422
	PPT_6 :	46/46	0/0
4	PPT_5 :	50/50	0/0
	PPT_6 :	59/59	1833/1833
5	PPT_3 :	202/202	0/0
	PPT_4 :	178/190	258/246
6	PPT_1 :	143/143	0/0
	PPT_2 :	55/83	327/244

Finally, for comparing the performance of the “Integrated” FDI scheme with the “Low Level” FDI scheme in [8] the results presented in Table 9 are evaluated by using the Confusion Matrix approach [20]. These results are shown in Table 10.

Table 10. Performance Results for Our Proposed “Low Level” and “Integrated” FDI Schemes

	“Low Level” FDI Scheme Performance Results	“Integrated” FDI Scheme Performance Results
Accuracy	95.16%	99.36%
True Healthy	100.00%	100.00%
False Healthy	0.08%	00.01%
True Faulty	92.48%	99.01%
False Faulty	0.00%	00.00%
Precision	87.99	98.24%

4 Conclusions

A novel Fault Detection and Isolation (FDI) scheme for Pulsed Plasma Thrusters (PPTs) of the Attitude Control Subsystem (ACS) of satellites in the formation flying missions is proposed and investigated. By means of four Dynamic Neural Networks (DNN) in each satellite the proposed FDI scheme is capable of detecting and isolating faults in the actuators (i.e. PPTs) of all the satellites which affect the precision and mission requirements for the formation flying attitudes.

Due to the fact that the force generated by this type of actuator cannot be measured, and due to the lack of precise mathematical models, the development of a fault diagnostic system for PPTs is not a trivial effort. In this paper, we have demonstrated that our proposed FDI scheme is not computationally intensive and is a reliable tool for detecting and isolating faulty PPTs.

The results obtained show a high level of accuracy (99.36%) and precision (98.24%) and the misclassification rate of the False Healthy and the False Faulty parameters that are quite negligible. Therefore, the applicability of our proposed DNN technique for solving fault diagnosis problems in a highly complex nonlinear system such as the formation flying systems has been demonstrated.

Formation Flying missions are beginning to gain popularity due to the number of advantages that they provide. A significant reduction in the amount of hours spent by the ground station personnel can be achieved by implementing our proposed DNN-based FDI schemes. Therefore, the cost of the missions can be reduced significantly.

References

1. Wilson, E., Lages, C., Mah, R.: Gyro-based Maximum-Likelihood Thruster Fault Detection and Identification. In: Proceedings of the 2002 American Control Conference (2002)
2. Wilson, E., Sutter, D.W., Berkovitz, D., Betts, B.J., del Mundo, R., Kong, E., Lages, C.R., Mah, R., Papasin, R.: Motion-based System Identification and Fault Detection and Isolation Technologies for Thruster Controlled Spacecraft. In: Proceedings of the JANNAF 3rd Modeling and Simulation Joint Subcommittee Meeting (2005)
3. Pirmoradi, F., Sassani, F., da Silva, C.W.: An Efficient Algorithm for Health Monitoring and Fault Diagnosis in a Spacecraft Attitude Determination System. In: IEEE International Conference on Systems, Man and Cybernetics (2007)
4. Larson, E.C., Parker Jr., B.E., Clark, B.R.: Model-Based Sensor and Actuator Fault Detection and Isolation. In: Proceedings of the American Control Conference (2002)
5. Joshi, A., Gavriloiu, V., Barua, A., Garabedian, A., Sinha, P., Khorasani, K.: Intelligent and Learning-based Approaches for Health Monitoring and Fault Diagnosis of RADAR-SAT-1 Attitude Control System. In: IEEE International Conference on Systems, Man and Cybernetics (2007)
6. Guiotto, Martelli, A., Paccagnini, C.: SMART-FDIR: Use of Artificial Intelligence in the Implementation of a Satellite FDIR. In: Data Systems in Aerospace DASIA 2003 (2003)
7. Holsti, N., Paakko, M.: Towards Advanced FDIR Components. In: DASIA 2001 Conference (2001)
8. Valdes, A., Khorasani, K.: Dynamic Neural Network-based Pulsed Plasma Thruster (PPT) Fault Detection and Isolation for the Attitude Control System of a Satellite. In: Proceedings of the 2008 International Joint Conference on Neural Networks (2008)

9. Pencil, E.J., Kamhawi, H., Arrington, L.A.: Overview of NASA's Pulsed Plasma Thruster Development Program. In: 40th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit (2004)
10. Bromaghim, D.R., Spanjers, G.G., Spores, R.A., Burton, R.L., Carroll, D., Schilling, J.H.: A Proposed On-Orbit Demonstration of an Advanced Pulsed-Plasma Thruster for Small Satellite Applications. Defense Technical Information Center OAI-PMH Repository (1998)
11. McGuire, M.L., Roger, Myers, M.: Pulsed Plasma Thrusters for Small Spacecraft Attitude Control. In: NASA/GSFC Flight Mechanics/Estimation Theory Symposium (1996)
12. Zakrzewski, C., Benson, S., Cassady, J., Sanneman, P.: Pulsed Plasma Thruster (PPT) Validation Report. NASA/GSFC (2002)
13. Li, Z.Q., Ma, L., Khorasani, K.: Dynamic Neural Network-Based Fault Diagnosis for Attitude Control Subsystem of a Satellite. In: Yang, Q., Webb, G. (eds.) PRICAI 2006. LNCS (LNAI), vol. 4099, pp. 308–318. Springer, Heidelberg (2006)
14. Al-Zyoud, I.A., Khorasani, K.: Neural Network-based Actuator Fault Diagnosis for Attitude Control Subsystem of an Unmanned Space Vehicle. In: International Joint Conference on Neural Networks (2006)
15. Korbicz, J., Obuchowicz, A., Patan, K.: Network of Dynamic Neuron in Fault Detection Systems. IEEE Systems, Man, and Cybernetics (1998)
16. Patan, K., Parisini, T.: Identification of Neural Dynamic Models for Fault Detection and Isolation: the Case of a Real Sugar Evaporation Process. *Journal of Process Control*, 67–79 (2005)
17. Ayoubi, M.: Nonlinear Dynamic Systems Identification with Dynamic Neural Networks for Fault Diagnosis in Technical Processes. In: IEEE International Conference on Systems, Man, and Cybernetics, vol. 3, pp. 2120–2125 (1994)
18. Patan, K.: Fault Detection of Actuators using Dynamic Neural Networks. In: 2nd Damadics Workshop on Neural Methods for Modeling and Fault Diagnosis (2003)
19. Computer Science 831: Knowledge Discovery in Databases,
http://www2.cs.uregina.ca/~dbd/cs831/notes/confusion_matrix/confusion_matrix.html

Passivity Analysis of a General Form of Recurrent Neural Network with Multiple Delays

Jinhua Huang and Jiqing Liu

Department of Electric and Electronic Engineering
Wuhan Institute of Shipbuilding Technology, Wuhan, Hubei, 430050, China
Angela_icec@yahoo.com.cn
LJQ6521@public.wh.hb.cn

Abstract. In this paper, by using some analytic techniques, several sufficient conditions are given to ensure the passivity of a general form of recurrent neural network with multiple delays. The passivity conditions are presented in terms of a negative semi-definite matrix declared. They are easily verifiable and easier to check computing with some conditions in terms of complicated linear matrix inequality.

Keywords: Passivity, Activation function, Multiple delays.

1 Introduction

Models of neural networks have recently attracted attention due to their promising potential for the tasks of pattern classification, designing associative memory, reconstruction of moving images, signal processing, parallel computation and solving some classes of optimization problems [1]. As is well known, most of these applications depend on their stability behavior of the neural networks. The problem of stability analysis of neural networks has been the central focus of numerous research activities. Some directions arise when dealing with typical applications or by placing constraint conditions on the network parameters of the neural system to ensure the desired stability properties. For example, when a neural network is designed to function as an associative memory [2], it is required that there exist several stable equilibrium points, whereas in the case of solving optimization problems, it is necessary that the designed neural networks must have a unique equilibrium point that is globally asymptotically stable. Therefore, it is of great interest to establish conditions that ensure the global asymptotic stability of a unique equilibrium point of a neural network.

In addition, within practical realizations of neural networks, the finite switching speed of amplifiers and active devices as well as the inherent communication time of neurons will incur time-delays in the interaction among the neurons. Time-delays, which are unavoidable in neural networks, may induce instability of the neural networks. Therefore, the stability analysis of delayed neural networks has been received much attention in recent years. It is concluded that delay effect might be the source of instability, hidden oscillations, divergence, chaos or other poor performance behavior. Delay-dependent stability conditions, which

contain information concerning time-delays, are usually less conservative than delay-independent ones, especially for a neural network with a small time-delay. As a result, recently, much attention has been paid on the delay-dependent stability analysis for delayed neural networks.

Usually, constant fixed time delays in the models of delayed feedback systems serve as good approximation in simple circuits having a small number of cells. Though delays arise frequently in practical applications, it is difficult to measure them precisely. In most situations, delays are variable, and in fact unbounded. That is, the entire history affects the present. Such delay terms, more suitable to practical neural nets, are called unbounded delays. Therefore, the studies of neural networks with time-varying delays and unbounded time delays are more important and actual than those with constant delays.

In the design of neural networks, the issue of global exponential stability is of prime concern since it guarantees the neural networks to converge fast enough in order to attain fast and satisfactory response. Accordingly, the problem of global exponential stability analysis for delayed neural networks has been studied by many investigators in the past years. In the case with with time-varying delays, sufficient conditions for global exponential stability were given in [3]-[13].

The notion of “passivity” of an input-output system, motivated by the dissipation of energy across resistors in an electrical circuit, has been widely used in order to analyze stability of a general class of interconnected nonlinear systems [14]. The passivity theory plays an important role in electrical networks and many other dynamical systems [15]. The passivity approach can also be used to neural networks. In [16], the authors have addressed the passivity properties of static multi-layer neural networks. Passivity analysis for dynamical multi-layer neuro identifier has been studied in [17], [18]. Recently, passivity analysis for delayed neural networks has been discussed by many researchers. Based on Lyapunov-Krasovskii theory, passivity conditions for neural networks with time-invariant delay and parametric uncertainties have been presented via LMIs [19]. The authors of [20] have extended the neural networks of [19] to integro-differential ones with time-varying delays. The delay-dependent results in [21] are less conservative than the delay-independent ones in [19] and [20]. In [3] and [23], the global dissipativity of neural networks have been examined based on internal state-space approach coupled with positive invariant sets. It must be noted that the concept of dissipativity is important in dynamical systems in general and in neural networks in particular and it has found applications in areas such as stability theory, chaos and synchronization theory, system norm estimation, and robust control. The global dissipativity of several classes of neural networks were discussed, and some sufficient conditions for the global dissipativity of neural networks with constant delays are derived in [22].

The remaining part of this paper consists of three sections. Section 2 describes some preliminaries. By using a negative semi-definite matrix, we get some sufficient conditions on passivity of a general form of recurrent neural network with multiple delays in Section 3. Finally, we make the conclusions in Section 4.

2 Preliminaries

Consider a general form of recurrent neural network

$$\begin{cases} \frac{dx_i(t)}{dt} = -d_i x_i(t) + \sum_{j=1}^n a_{ij} f_j(x_j(t)) + \sum_{j=1}^n b_{ij} f_j(x_j(t - \tau_j)) + u_i(t), \\ y(t) = f(x(t)), \end{cases} \tag{1}$$

where $i = 1, 2, \dots, n, d_i > 0$ is positive parameter, $u_i(t)$ is the input, $u(t) = (u_1(t), \dots, u_n(t))^T$, $x_i(t)$ is the state of the i th neuron, $x(t) = (x_1(t), \dots, x_n(t))^T$, $A = (a_{ij})_{n \times n}, B = (b_{ij})_{n \times n}$ are the connection weight matrices that are not assumed to be symmetric, and $f(x(t)) = (f_1(x_1(t)), f_2(x_2(t)), \dots, f_n(x_n(t)))^T$ is the activation function. Denote I as an n -dimensional identity matrix. Let

$$\delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

Definition 1. The neural network model (1) is called passive if there exists a scalar $\gamma \geq 0$ such that

$$2 \int_0^{t_p} y^T(s)u(s)ds \geq -\gamma \int_0^{t_p} u^T(s)u(s)ds \tag{2}$$

for all $t_p \geq 0$ and for all solutions of (1) with $x_0 = 0$.

3 Main Results

Theorem 1. Let the continuous function $f(\cdot)$ belong to the following set:

$$\left\{ f(\cdot) \mid 0 \leq \frac{f_i(r)}{r} \leq \ell_i < \infty, \forall r \in \mathfrak{R}, i = 1, 2, \dots, n \right\}. \tag{3}$$

If the following matrix Q_1 is negative semi-definite, then the neural network model (1) is passive, where

$$Q_1 = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{12}^T & Q_{22} \end{pmatrix},$$

$$Q_{11} = 2(a_{ij} + \delta_{ij}(\frac{1}{2} - \frac{d_i}{\ell_i}))_{n \times n}, Q_{12} = (b_{ij})_{n \times n}, Q_{22} = -I_{n \times n}.$$

Proof. We employ the following function:

$$V(x(t), t) = 2 \sum_{i=1}^n \int_0^{x_i} f_i(x_i)dx_i + \sum_{i=1}^n \int_{t-\tau_i}^t f_i^2(x_i(\xi))d\xi. \tag{4}$$

Computing the derivative of $V(x(t), t)$ along the positive half trajectory of (1), we have

$$\begin{aligned}
 & \frac{dV(x(t), t)}{dt} \Big|_{\text{II}} - 2y^T(t)u(t) - \gamma u^T(t)u(t) \\
 &= 2 \sum_{i=1}^n \left[-d_i x_i(t) f_i(x_i(t)) + \sum_{j=1}^n a_{ij} f_i(x_i(t)) f_j(x_j(t)) \right. \\
 & \quad \left. + \sum_{j=1}^n b_{ij} f_i(x_i(t)) f_j(x_j(t - \tau_j)) + f_i(x_i(t)) u_i(t) \right] \\
 & \quad + \sum_{i=1}^n f_i^2(x_i(t)) - \sum_{i=1}^n f_i^2(x_i(t - \tau_i)) - 2y^T(t)u(t) - \gamma u^T(t)u(t) \\
 &= 2 \sum_{i=1}^n \left[-d_i x_i(t) f_i(x_i(t)) + \sum_{j=1}^n a_{ij} f_i(x_i(t)) f_j(x_j(t)) \right. \\
 & \quad \left. + \sum_{j=1}^n b_{ij} f_i(x_i(t)) f_j(x_j(t - \tau_j)) \right] + \sum_{i=1}^n f_i^2(x_i(t)) \\
 & \quad - \sum_{i=1}^n f_i^2(x_i(t - \tau_i)) - \gamma u^T(t)u(t) \\
 &\leq 2 \sum_{i=1}^n \left[\left(\frac{1}{2} - \frac{d_i}{\ell_i} \right) f_i^2(x_i(t)) + \sum_{j=1}^n a_{ij} f_i(x_i(t)) f_j(x_j(t)) \right. \\
 & \quad \left. + \sum_{j=1}^n b_{ij} f_i(x_i(t)) f_j(x_j(t - \tau_j)) \right] - \sum_{i=1}^n f_i^2(x_i(t - \tau_i)) - \gamma u^T(t)u(t) \\
 &= 2 \sum_{i=1}^n \left[\sum_{j=1}^n \left(a_{ij} + \delta_{ij} \left(\frac{1}{2} - \frac{d_i}{\ell_i} \right) \right) f_i(x_i(t)) f_j(x_j(t)) \right. \\
 & \quad \left. + \sum_{j=1}^n b_{ij} f_i(x_i(t)) f_j(x_j(t - \tau_j)) \right] - \sum_{i=1}^n f_i^2(x_i(t - \tau_i)) - \gamma u^T(t)u(t) \\
 &= \begin{pmatrix} f(x(t)) \\ f(x(t - \tau)) \\ u(t) \end{pmatrix}^T \begin{pmatrix} Q_1 & 0 \\ 0 & -\gamma \end{pmatrix} \begin{pmatrix} f(x(t)) \\ f(x(t - \tau)) \\ u(t) \end{pmatrix} \\
 &\leq 0;
 \end{aligned}$$

i.e.,

$$\frac{dV(x(t), t)}{dt} \Big|_{\text{II}} - 2y^T(t)u(t) - \gamma u^T(t)u(t) \leq 0. \tag{5}$$

By integrating (5) with respect to t over the time period $[0, t_p]$,

$$2 \int_0^{t_p} y^T(s)u(s)ds \geq V(x(t_p), t_p) - V(x(t_0), t_0) - \gamma \int_0^{t_p} u^T(s)u(s)ds. \tag{6}$$

For $x_0 = 0$, we have $V(x(t_0), t_0) = 0$. Hence, according to Definition 1, (II) is passive.

Theorem 2. Let the function $f(\cdot)$ belong to the following set:

$$\left\{ f(\cdot) \mid f_i(x) \in C[\mathbb{R}, \mathbb{R}], \quad D^+ f_i(x_i) \geq 0, \quad i = 1, 2, \dots, n \right\}. \tag{7}$$

If the following matrix

$$Q_2 = \begin{pmatrix} Q_{11} & Q_{12} & Q_{13} & I_{n \times n} \\ Q_{12}^T & Q_{22} & Q_{23} & 0 \\ Q_{13}^T & Q_{23}^T & 0 & 0 \\ I_{n \times n} & 0 & 0 & -\gamma \end{pmatrix}$$

is negative semi-definite, where $Q_{11} = -2diag(d_i)_{n \times n}$, $Q_{12} = (a_{ij} - \delta_{ij}d_i)_{n \times n}$, $Q_{13} = Q_{23} = (b_{ij})_{n \times n}$, $Q_{22} = 2(a_{ij})_{n \times n}$, then the neural network model (II) is passive.

Proof. We employ the following function:

$$V(x(t), t) = \sum_{i=1}^n x_i^2(t) + 2 \sum_{i=1}^n \int_0^{x_i} f_i(x_i) dx_i. \tag{8}$$

Computing the derivative of $V(x(t), t)$ along the trajectory of (II), we have

$$\begin{aligned} & \frac{dV(x(t), t)}{dt} \Big|_{(II)} - 2y^T(t)u(t) - \gamma u^T(t)u(t) \\ &= 2 \sum_{i=1}^n \left[-d_i x_i^2(t) - d_i x_i(t) f_i(x_i(t)) \right. \\ & \quad + \sum_{j=1}^n a_{ij} x_i(t) f_j(x_j(t)) + \sum_{j=1}^n a_{ij} f_i(x_i(t)) f_j(x_j(t)) \\ & \quad + \sum_{j=1}^n b_{ij} x_i(t) f_j(x_j(t - \tau_j)) + \sum_{j=1}^n b_{ij} f_i(x_i(t)) f_j(x_j(t - \tau_j)) \\ & \quad \left. + x_i(t) u_i(t) + f_i(x_i(t)) u_i(t) \right] - 2y^T(t)u(t) - \gamma u^T(t)u(t) \\ &= 2 \sum_{i=1}^n \left[-d_i x_i^2(t) - d_i x_i(t) f_i(x_i(t)) + \sum_{j=1}^n a_{ij} x_i(t) f_j(x_j(t)) \right. \\ & \quad + \sum_{j=1}^n a_{ij} f_i(x_i(t)) f_j(x_j(t)) + \sum_{j=1}^n b_{ij} x_i(t) f_j(x_j(t - \tau_j)) \\ & \quad \left. + \sum_{j=1}^n b_{ij} f_i(x_i(t)) f_j(x_j(t - \tau_j)) + x_i(t) u_i(t) \right] - \gamma u^T(t)u(t) \\ &\leq \begin{pmatrix} x(t) \\ f(x(t)) \\ f(x(t - \tau)) \\ u(t) \end{pmatrix}^T Q_2 \begin{pmatrix} x(t) \\ f(x(t)) \\ f(x(t - \tau)) \\ u(t) \end{pmatrix} \leq 0; \end{aligned}$$

i.e.,

$$\frac{dV(x(t), t)}{dt} \Big|_{\text{(9)}} - 2y^T(t)u(t) - \gamma u^T(t)u(t) \leq 0. \quad (9)$$

By integrating (9) with respect to t over the time period $[0, t_p]$,

$$2 \int_0^{t_p} y^T(s)u(s)ds \geq V(x(t_p), t_p) - V(x(t_0), t_0) - \gamma \int_0^{t_p} u^T(s)u(s)ds. \quad (10)$$

For $x_0 = 0$, we have $V(x(t_0), t_0) = 0$. Hence, according to Definition 1, (11) is passive.

4 Concluding Remarks

This paper presents two sufficient conditions on passivity for a general form of recurrent neural network. The passivity conditions are presented in terms of a negative semi-definite matrix. These conditions are very easy to be verified, and useful to analysis and characterize the passivity of the neural networks.

References

1. Forti, M., Tesi, A.: New Conditions for Global Stability of Neural Networks with Application to Linear and Quadratic Programming Problems. *IEEE Trans. Circ. Syst.* 42, 354–366 (1995)
2. Zeng, Z.G., Wang, J.: Analysis and Design of Associative Memories Based on Recurrent Neural Networks with Linear Saturation Activation Functions and Time-varying Delays. *Neural Computation* 19, 2149–2182 (2007)
3. Liao, X.X., Wang, J.: Algebraic Criteria for Global Exponential Stability of Cellular Neural Networks with Multiple Time Delays. *IEEE Trans. Circuits and Systems* 50, 268–275 (2003)
4. Yi, Z., Heng, A., Leung, K.S.: Convergence Analysis of Cellular Neural Networks with Unbounded Delay. *IEEE Trans. Circuits Syst.* 48, 680–687 (2001)
5. Chen, T.P., Rong, L.B.: Robust Global Exponential Stability of Cohen-Grossberg Neural Networks with Time-Delays. *IEEE Transactions on Neural Networks* 15, 203–206 (2004)
6. Liao, X.F., Li, C.G., Wong, K.W.: Criteria for Exponential Stability of Cohen-Grossberg Neural Networks. *Neural Networks* 17, 1401–1414 (2004)
7. Cao, J.: Results Concerning Exponential Stability and Periodic Solutions of Delayed Cellular Neural Networks. *Physics Letters* 307, 136–147 (2003)
8. Zeng, Z.G., Wang, J., Liao, X.X.: Global Exponential Stability of A General Class of Recurrent Neural Networks with Time-varying Delays. *IEEE Trans. Circuits and Systems Part* 50, 1353–1358 (2003)
9. Zeng, Z.G., Wang, J., Liao, X.X.: Stability Analysis of Delayed Cellular Neural Networks Described Using Cloning Templates. *IEEE Trans. Circuits and Syst.* 51, 2313–2324 (2004)
10. Zeng, Z.G., Wang, J.: Improved Conditions for Global Exponential Stability of Recurrent Neural Network with Time-varying Delays. *IEEE Trans on Neural Networks* 17, 623–635 (2006)

11. Zeng, Z.G., Wang, J.: Global Exponential Stability of Recurrent Neural Networks with Time-varying Delays in the Presence of Strong External Stimuli. *Neural Networks* 19, 1528–1537 (2006)
12. Zeng, Z.G., Wang, J.: Multiperiodicity of Discrete-time Delayed Neural Networks Evoked by Periodic External Inputs. *IEEE Transactions on Neural Networks* 17, 1141–1151 (2004)
13. Zeng, Z.G., Wang, J.: Complete Stability of Cellular Neural Networks with Time-varying Delays. *IEEE Transactions on Circuits and Systems* 53, 944–955 (2006)
14. Byrnes, C.I., Isidori, A., Willems, J.C.: Passivity, Feedback Equivalence, and the Global Stabilization of Minimum Phase Nonlinear Systems. *IEEE Transactions on Automatic Control* 36, 1228–1240 (1991)
15. Lozano, R., Brogliato, B., Egeland, O., Maschke, B.: *Dissipative Systems Analysis and Control: Theory and Applications*. Springer, London (2000)
16. Commuri, S., Lewis, F.L.: CMAC Neural Networks for Control of Nonlinear Dynamical Systems: Structure, Stability, and Passivity. *Automatica* 33, 635–641 (1997)
17. Yu, W., Li, X.: New Results on System Identification with Dynamic Neural Networks. *IEEE Trans. Neural Networks* 12, 412–417 (2001)
18. Yu, W.: Passivity Analysis for Dynamic Multilayer Neuro Identifier. *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.* 50, 173–178 (2003)
19. Li, C., Liao, X.: Passivity Analysis of Neural Networks with Time Delays. *IEEE Trans. Circuits Syst. II, Exp. Briefs* 52, 471–475 (2005)
20. Lou, X., Cui, B.: Passivity Analysis of Integro-differential Neural Networks with Time-varying Delays. *Neurocomputing* 70, 1071–1078 (2007)
21. Park, J.H.: Further Results on Passivity Analysis of Delayed Cellular Neural Networks. *Chaos, Solitons and Fractals* 34, 1546–1551 (2007)
22. Liao, X.X., Wang, J.: Global Dissipativity of Continuous-Time Recurrent Neural Networks with Time Delay. *Phys. Rev.* 68, 1–7 (2003)
23. Song, Q., Zhao, Z.: Global Dissipativity of Neural Networks with both Variable and Unbounded Delays. *Chaos, Solitons and Fractals* 25, 393–401 (2005)

Comparative Analysis of Corporate Failure Prediction Methods: Evidence from Chinese Firms

Haicong Yang

Hubei University of Economics, 430000 Wuhan, China
hcyang2002@163.com

Abstract. This paper examines four most popular alternative methods that have been applied in financial failure prediction: linear discriminant analysis, logit analysis, neural networks and support vector machine. The main purpose is to make comparisons of the prediction abilities among different methods. It was implemented by using the Chinese firms data one, two and three years prior to failure in the empirical analysis. The results indicate that the classification accuracy of support vector machine is the highest.

Keywords: Failure prediction, Statistical method, Intelligent method.

1 Introduction

Many studies have been conducted since Beaver [1] firstly used financial ratios in failure prediction. The studies can be classified into three categories: (1) Theoretical modelling of the failure process; (2) empirical researches on failure forecasting variables, which contain financial and nonfinancial variables; (3) researches for the most effective empirical forecasting methods. The first two aspects haven't got breakthrough findings, whereas the research on the third aspect has generated continual new achievements. All these studies have focused on the effect of the violation of the assumptions set by the methods, the effect of the statistical samples and the development or application of new methods.

The purpose of applying new methods is to increase the accuracy of failure prediction. Linear discriminant analysis (LDA) was first applied in failure prediction in the 1960s[2]. LDA was replaced by logit analysis in the 1970s and 1980s[3]. Neural networks (NN) have been introduced in failure prediction in 1990s [4]. Since the beginning of 21 century, the support vector machine (SVM) based on the small sample learning theory has been applied to solve classification problems.

Recently, Chinese economy attracts many attentions and Chinese economy obviously belongs to a transition economy. Chinese listed firms show different characteristics from those in developed market economy. However, till now almost all researches on the financial distress classification of Chinese corporates have not made a complete comparison for different classification tools, but just applied different tools to different samples[5][6]. This study tries to conduct the empirical comparison among LDA, Logit analysis, NN and SVM.

The following sections are organized as follows: section 2 briefly introduces the selected failure prediction methods; section 3 is about the sample and variables that were used in the research; section 4 is about the effect of the ex-post classification and ex-ante prediction of the selected methods; and the last section concludes the study.

2 Methods on Failure Prediction

2.1 Linear Discriminant Analysis (LDA)

The purpose of LDA is to discriminate among the groups by a linear combination of predictors. A score is calculated based on the set of independent predictors for each firm, and then a cut-off point is established so that the firms with a score below the cut-off point are expected to fail while those with a score above the point are expected to be going-concern. The score can be calculated as follows:

$$Z = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_n X_n \quad (1)$$

Where X_i ($i = 1, \dots, n$) are the independent variables and b_i ($i = 1, \dots, n$) are the estimated parameters.

The use of LDA is restricted by two statistical requirements. First, the independent variables should be multivariate normal, and second, the covariance matrices of the two groups should be the same. Both assumptions are often violated because of the financial ratios which are selected as predictors can not meet the requirements.

2.2 Logit Analysis (Logit)

Logit analysis method is also used a score to decide whether a firm belongs to a group (failed or non-failed) or not, the failure probability is calculated as follows:

$$P(Z) = 1/(1 + \exp(-Z)) = 1/(1 + \exp(-(b_0 + b_1 X_1 + b_2 X_2 + \dots + b_n X_n))) \quad (2)$$

Where X_i ($i = 1, 2, \dots, n$) are the independent variables and b_i ($i = 1, 2, \dots, n$) are the estimated parameters. The output $P(z)$ is a hyperbolic function, and its value is between 0 and 1.

The result of $P(Z)$ can be regarded as the conditional probability of failure. If the selected cumulative distribution function (CDF) is normal cumulative distribution function other than logistic function under the binomial model, it is then called as probit model. It is difficult to estimate the coefficient because the model is non-linear, so, the coefficients are always estimated by maximum likelihood estimation. It generates the probability of a group membership since its value changes between 0 and 1. If $Z \rightarrow -\infty$, then $P(Z) \rightarrow 0$; while $Z \rightarrow +\infty$, then $P(Z) \rightarrow 1$; if $Z = 0$, then $P(Z) = 0.5$, which is a commonly used critical value in classifying failed and

non-failed firms. Type I error means that failed firms are classified as non-failed. Type II errors means that non-failed firms are classified as failed. If misclassification costs for both error types are taken into account (misclassification costs for the Type I error are usually estimated to be higher than those of the Type II error, because type I error will make the investors occur actual loss while type II only causes opportunity costs). The midranges of probability are more sensitive to the changes in the values of independent variables.

2.3 Neural Networks (NN)

In 1990s, a new method named neural networks was broadly applied in the model identification area, in which include classification problems process. Widrow and Hoff [7] described the first neural-based computer. Hopfield showed that there are many problems that could be solved by using NN [8].

Before an NN is used to predict it must be trained or learned by using a group of observations, i.e. a group of failed and non-failed firms. This can be supervised or unsupervised. In the former case the outcome of every observation is known and the network trains itself until it can combine a certain input with a certain outcome. When unsupervised learning is used, the outcomes of observations are not given and NN self-organizes the input data and discovers the basic features of it. Thus, NN learns independently to associate these basic features with a certain outcome.

After the training has been completed, the NN can be used in prediction. The structure of network, include the neurons and interconnections between them, is now fixed. The training and the whole NN can be evaluated on the basis of the number of correct predictions. The more correct predictions the trained NN makes, the more successful it is.

2.4 Support Vector Machine(SVM)

Support Vector Machine is developed based on Statistical Learning Theory [9] which focuses on machine learning law under the condition of small sample. Along with the development of the theory, and also because of the material lack of progress of other approaches such as NN, SVM method came into being since the middle of 1990s, and shows its advantages over the existed methods.

The underlying theme of the class of supervised learning methods is to learn from observations. There is an input space, denoted by $X \subseteq R^n$, an output space, denoted by Y , and a training set, denoted by S , $S = ((x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)) \subseteq (X \times Y)^l$, l is the size of the training set. The overall assumption for learning is the existence of a hidden function $Y = f(X)$, and the task of classification is to construct a heuristic function $h(X)$, such that $h \rightarrow f$ on the prediction of Y . The nature of the output space Y decides the learning type. $Y = \{-1, 1\}$ leads to a binary classification problem,

$Y = \{1, 2, \dots, m\}$ leads to a multiple class classification problem, and $Y \subseteq R^n$ leads to a regression problem.

For binary classification problem, when classification problem belongs to linearly separable, the final SVM classifier is as follows:

$$Y = \text{sign}\left(\sum_{i=1}^N y_i \alpha_i \langle x, x_i \rangle + b\right) \quad (3)$$

For binary classification problem, when classification problem belongs to linearly non-separable, the final SVM classifier is as follows:

$$Y = \text{sign}\left(\sum_{i=1}^N y_i \alpha_i K(x, x_i) + b\right) \quad (4)$$

Where $K(x, x_i)$ is kernel function, $K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$, it transforms the computations of $\langle \phi(x_i), \phi(x_j) \rangle$ to that of $\langle x_i, x_j \rangle$. $\langle \cdot, \cdot \rangle$ represents inner product. $\phi(x)$ represents high-dimensional feature space which is nonlinearly mapped from the input space x . N is the number of support vectors which are data instances corresponding to non-zero α_i 's. b and α_i are the coefficients to be estimated. For more details about estimation process please refer to Vapnik's monograph [9].

3 Research Design

3.1 Sample Selection

The sample was selected from the companies in Shanghai and Shenzhen stock exchange market. Two groups of samples were used: (1) failed group; (2) random group. The failed firms are those being announced to be delisted because of a 3-continuous-year's loss during 1999-2005. The selected failed firms have complete and available financial data of 3 years prior to bankruptcy. Financial firms were excluded because of the feature of the financial reports. The model that uses financial ratios cannot mix the financial and non-financial firms together. After exclusion, 80 failed firms were left.

The random group included 80 firms is selected from non-financial firms with complete data other than failed group. Each failed firm is matched with a non-failed one randomly with the same period to ensure the data proportion of the failed and non-failed firms is the same.

The above-mentioned two groups formed the failed/ non-failed sample of this research. Each group is divided by the proportion of 7:3 into estimation sample and test sample, the former includes 56 failed firms and 56 random firms, the latter includes 24 failed firms and 24 random firms, the firms allocated to the two subsamples are selected at random.

3.2 Prediction Variables

The purpose of this research is not to increase the accuracy of failure prediction, but to make comparisons of the prediction abilities among different methods, so, the number of the samples is small to a minimum to simplify the research. The study uses leverage ratio (measured by the ratio of total debts to total assets, simplified as TDTA) as the proxy variable of the failure because the more debt a firm has, the higher probability the firm has. The study uses liquidity ratio (current assets to current liabilities, CACL) as a proxy variable of the failure because a firm will meet operational difficult, and cannot repay the due debt. The study uses operating income to total assets (OITA) as a proxy of failure because profitability is the premise of continuum operation. This research also pays enough attention to the special phenomenon of substantial shareholders encroach the funds of published firms in China, and takes it as a prediction variable. This paper uses receivables to owners' equity as the proxy of the situation.

3.3 Dependent Variable

The financial states used in the study are: (1) state 0: financial health, without financial distress events; (2) state 1: receiving delisted treatment. Dependent variable in this research can achieve two values and is coded as follows:

STAT= 0, if the firm is health;

1, if the firm receives delisted treatment.

4 Empirical Results

4.1 Results of Ex-post Classification

Four methods were used to conduct the discriminate classification on the estimated data in order to compare the classification results. In the ex-post classification, the accuracy of discriminate classifications for 1, 2 and 3 year prior to failure are 85.7%, 71.4% and 72.4% respectively, there is no obvious difference in the number of the type I and II error at the 1st year prior to failure, but as to the 2nd and 3rd year prior to failure, the type II error is predominant, that is, the rate of misclassifying non-failed firms is more than that of failed firms. (as shown in table 1)

The result of Logit analysis on the 112 samples shows that the number of type I and II error of the 1st year prior to failure are both 5, the classification accuracy is 91.1%, the classification accuracy of the 2nd and 3rd year prior to failure is 72.3% and 75.9% respectively.

Back propagation algorithm is used to establish and train the neural networks model of the 1st, 2nd and 3rd year prior to failure. In this research, there are 3 neurons for each network, 5-15 implied neurons and 1 output neuron. In training stage, there are 9 firms

among the 112 samples of the 1st year prior failure are misclassified, that is, the rate of misclassification is 9%, it increase to 12.5% in the 2nd year prior to failure, and 12.5% in the 3rd year. The total misclassification rate in the 1st, 2nd and 3rd year prior to failure is 8%, 11.6% and 12.5% respectively. So, the discriminant effect of SVM is relatively ideal.

Table 1. Comparison of the ex-post classification results

	1st year			2nd year			3rd year		
	Type I	Type II	Total	Type I	Type II	Total	Type I	Type II	Total
LDA	14.3	14.3	14.3	26.8	28.6	28.6	23.2	32.1	27.6
	%	%	%	%	%	%	%	%	%
Logit	8.90	8.90	8.90	26.80	28.60	27.70	23.20	32.10	24.10
	%	%	%	%	%	%	%	%	%
NN	7.10	8.90	8.00	10.70	14.30	12.50	17.90	7.10	12.50
	%	%	%	%	%	%	%	%	%
SVM	7.10	8.90	8.00	8.90	14.30	11.60	16.10	8.90	12.50
	%	%	%	%	%	%	%	%	%

Table 2. Methods ranked ex ante prediction accuracy

1st year		2nd year		3rd year	
I	Neural networks	I	SVM	I	SVM
I	SVM	II	Neural networks	I	Neural networks
II	Logit	III	Logit	II	Logit
III	LDA	IV	LDA	III	LDA

Table 2 shows the results of classification accuracy ordering under each method. The misclassification result of LDA is the highest in each year, while it is the lowest under neural networks. The results illustrate that the number of correct classification is interrelated with the used method.

4.2 Ex-ante Prediction Results

After the model is estimated, the prediction ability of the model can be tested with the test set. The results were shown in table 3. The SVM method was the best in the 1st year prior to failure, the prediction accuracy was as high as 89.6%, and that of neural networks and logit analysis were both 85.4%.

As to the 2nd year prior to failure, the result obviously went worse, the accuracy of neural networks and SVM were 72.9%, while that of logit analysis and LDA was

66.7% and 64.6% respectively. As to the 3rd year prior to failure, the accuracy of SVM was 81.2%, that of neural networks was 72.9%, and that of logit analysis and LDA were both 62.5%.

In table 4 the methods are ranked according to the ex-ante prediction accuracy, the results show that the prediction effect of SVM is the best, and that of the traditional statistical method such as Logit analysis and LDA is worse, among which LDA is the worst. The results also show that the application prospect of intelligent tools in classification research field is promising.

Table 3. Comparison of the ex-post classification results

	1st year			2nd year			3rd year		
	Type I	Type II	Total	Type I	Type II	Total	Type I	Type II	Total
LD	16.7	16.7	16.7	33.3	37.50	35.4	33.30%	41.70	37.50
A	0%	0%	0%	0%	%	0%		%	%
Lo	16.7	12.5	14.6	33.3	33.30	33.3	33.30%	41.70	37.50
git	0%	0%	0%	0%	%	0%		%	%
N	20.8	8.30	14.6	33.3	20.80	27.1	37.50%	16.70	27.10
N	0%	%	0%	0%	%	0%		%	%
SV	12.5	8.30	10.4	33.3	20.80	27.1	25%	12.50	18.80
M	0%	%	0%	0%	%	0%		%	%

Table 4. Methods ranked ex-ante prediction accuracy

1st year		2nd year		3rd year	
I	SVM	I	SVM	I	SVM
II	Neural networks	I	Neural networks	II	Neural networks
II	logit	II	logit	III	Logit
III	LDA	III	LDA	IV	LDA

5 Conclusions

To increase the prediction accuracy is one of the most important issues in failure prediction, and many methods of improving the prediction accuracy come up. This research focuses on the classification effect of LDA, Logit analysis, neural networks and SVM, use leverage ratio, liquidity ratio, profitability ratio and the occupation of the firm funds as the prediction variables, the results show that the classification effect of the intellectual method is better than traditional statistical method under the results of both ex-ante and ex-post circumstances, among which the effect of SVM is the best while that of LDA is the worst.

References

1. Beaver, W.: Alternative Accounting Measures as Predictors of Failure. *The Accounting Review* 4, 79–111 (1966)
2. Altman, E.I.: Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance* 9, 589–609 (1968)
3. Ohlson, J.: Financial Ratio and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research* 18, 109–131 (1980)
4. Tam, K., Kiang, M.: Managerial Application of Neural Networks: The Case of Bank Failure Prediction. *Management Science* 38, 926–947 (1992)
5. Wu, S.N., Lu, X.Y.: The Financial Distress Prediction Research of Chinese Listed Corporations. *Economic Research Journal* 6, 46–55 (2001)
6. Yang, S., Huang, L.: Firms Warning Model Based on BP Neural Networks. *Systems Engineering Theory and Practice* 1, 12–18 (2005)
7. Widrow, B., Hoff, M.: Adaptive Switching Circuits. *IRE-WESCON Convention Record* 4, 96–104 (1960)
8. Hopfield, J.: Neural Networks and Physical Systems with Emergent Collective Computational Abilities. *Proceedings of the National Academy of Science* 79, 2554–2558 (1982)
9. Vapnik, V.: *Statistical Learning Theory*. Springer, Heidelberg (1998)

An Adaline-Based Location Algorithm for Wireless Sensor Network

Fengjun Shang

College of Computer Science and Technology,
Chongqing University of Posts and Telecommunications, Chongqing 400065, China
shangfj@cqupt.edu.cn

Abstract. Localization is used in location-aware applications such as navigation, autonomous robotic movement, and asset tracking to position a moving object on a coordinate system. In this paper, we present an improved DV-Hop algorithm based on Adaline (adaptive linear neuron) and DV-Hop, called ADV-Hop. The algorithm makes three major contributions to the localization problem in the wireless sensor networks (WSNs). First, we present a practical, fast and easy-to-use localization scheme with relatively high accuracy and low cost for WSNs. Second, the proposed algorithm improves location accuracy than the DV-Hop algorithm. Third, we explored the influence of anchor nodes on localization performance of the ADV-Hop algorithm. The proposed method can improve location accuracy and coverage without increasing hardware cost of sensor node. Simulation results show that the performance of this algorithm is superior to the original DV-Hop algorithm. Compared with DV-Hop, it is more available for WSNs.

Key words: WSN, DV-Hop, Adaline, Location accuracy.

1 Introduction

With the development of sensor techniques, low-power electronic and radio techniques, low-power and inexpensive wireless sensors have been put into application, then the wireless sensor networks (WSNs) have appeared. WSNs can be applied to many areas, such as military affairs, commerce, medical care, environmental monitoring, and have become a new research focus in computer and communication fields. Many applications of WSNs are based on sensor self-positioning, such as battlefield surveillance, environment monitoring, indoor user tracking and others, which depend on knowing the location of sensor nodes. Because of the constraint in size, power, and cost of sensor nodes, the investigation of efficient location algorithms which satisfy the basic accuracy requirement for WSNs meets new challenges.

In cellular location estimation [1]–[3] and local positioning systems (LPS) [4], [5], location estimates are made using only ranges between a blindfolded device and reference devices. Relative location estimation requires simultaneous estimation of multiple device coordinates. Greater location estimation accuracy can be achieved as devices are added into the network, even when new devices have no a prior coordinate information and range to just a few neighbors.

Many localization algorithms for sensor networks have been proposed to provide per-node location information. Based on the type of knowledge used in localization, we divide these localization protocols into two categories: range-based and range-free. Range-based protocols use absolute point-to-point distance or angle information to calculate the location between neighboring sensors. The second class of methods, range-free approach, employs to find the distances from the non-anchor nodes to the anchor nodes. Several ranging techniques are possible for range measurement, such as angle-of-arrival (AOA) [6], received signal strength indicator (RSSI) [7], time-of-arrival (TOA) [8] or time-difference-of-arrival (TDOA)[9]. Because of the hardware limitations of WSNs devices, solutions in range-free localization are being pursued as a cost-effective alternative to more expensive range-based approaches [10]. Because of the advantages on power and cost on sensor node, this paper focuses the investigation on the range-free algorithms for WSNs [11][12].

Centroid algorithm [13] is a simple range-free localization algorithm. The node receives signals of landmarks in its communication area and makes its coordinates as the centroid of these landmarks. Additional devices of localization are not required in this algorithm. Thereby, the hardware of nodes can be simple, but its precision is comparatively low.

In this paper, we present an improved DV-Hop algorithm based on Adaline and DV-Hop. The proposed method can improve location accuracy without increasing hardware cost of sensor node. Simulation results show that the performance of this algorithm is superior to the original DV-Hop algorithm. Compared with DV-Hop, it is more available for WSNs.

This paper makes three major contributions to the localization problem in WSNs. First, we present a practical, fast and easy-to-use localization scheme with relatively high accuracy and low cost for WSNs. Second, the proposed algorithm improves location accuracy than the DV-Hop algorithm. Third, we explored the influence of anchor nodes on localization performance of the ADV-Hop algorithm.

The rest of this paper is organized as follows. Section 2 presents the derivation of the proposed improved DV-Hop algorithm. In Section 3, simulation results are shown and localization performances are discussed. Finally, we present our conclusions in Section 4.

2 Improved DV-Hop Location Scheme

Niculescu and Nath [14] have proposed the DV-Hop, which is a distributed, hop by hop positioning algorithm. The algorithm implementation is comprised of three steps. First, it employs a classical distance vector exchange so that all nodes in the network get distances, in hops, to the landmarks. And then, it estimates an average size for one hop, which is then deployed as a correction to the entire network. Finally, unknown nodes compute their location by trilateration [15].

2.1 DV-Hop Algorithm

In the first step, each anchor node broadcasts a beacon to be flooded throughout the network containing the anchors location with a hop-count value initialized to one. Each receiving node maintains the minimum hop-count value per anchor of all beacons it

receives. Beacons with higher hop-count values to a particular anchor are defined as stale information and will be ignored. Then those not stale beacons are flooded outward with hop-count values incremented at every intermediate hop. Through this mechanism, all nodes in the network get the minimal hop-count to every anchor node.

In the second step, once an anchor gets hop-count value to other anchors, it estimates an average size for one hop, which is then flooded to the entire network. After receiving hop-size, blindfolded nodes multiply the hop-size by the hop-count value to derive the physical distance to the anchor. The average hop-size is estimated by anchor i using the following formula:

$$HopDistance = \frac{\sum_{i \neq j} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}{\sum_{i \neq j} H_{ij}} \tag{1}$$

where $(x_i, y_i), (x_j, y_j)$ are coordinates of anchor i and anchor j , H_{ij} is the hops between beacon i and beacon j .

In the third phase, the unknown node locations can be estimated by the multilateration method when these nodes have the distance estimations to at least three reference nodes in the plane. Given a set of reference nodes $R_i = (x_i, y_i)^T, i.1, 2, \dots, M$, where M is the number of reference nodes, let the hop value between the unknown node $X = (x, y)^T$ and the i -th reference node is L_i . Then the distance between the unknown node and i -th reference node is given by $d_i = L_i \times HopDistance$. The unknown node location X can be obtained by

$$\begin{cases} (x_1 - x)^2 + (y_1 - y)^2 = d_1^2 \\ \dots \\ (x_M - x)^2 + (y_M - y)^2 = d_M^2 \end{cases} \tag{2}$$

In the above data structure, (x_i, y_i) are the two-dimensional coordinates of the i -th reference point, (x, y) are the coordinates of unknown node, and d_i is the measured ranged between the i -th reference point and the unknown. This data structure can be linearized by subtracting the last row and performing some minor arithmetic shuffling, resulting in the following relations:

$$AX = b \tag{3}$$

$$A = \begin{bmatrix} (x_1 - x_3) & (y_1 - y_3) \\ (x_2 - x_3) & (y_2 - y_3) \end{bmatrix} \tag{4}$$

$$X = \begin{bmatrix} x \\ y \end{bmatrix} \tag{5}$$

$$b = \begin{bmatrix} d_1^2 - d_3^2 - x_1^2 + x_3^2 - y_1^2 - y_3^2 \\ d_2^2 - d_3^2 - x_2^2 + x_3^2 - y_2^2 - y_3^2 \end{bmatrix} \tag{6}$$

The above set of equations can be solved using a traditional least squares algorithm and the least squares estimate of X is written as

$$\hat{X} = (A^T A)^{-1} A^T b \tag{7}$$

Each anchor node broadcasts its *HopDistance* to network using controlled flooding. Unknown nodes receive *HopDistance* information, and save the first one. At the same time, they transmit the *HopDistance* to their neighbor nodes. This scheme could assure that the most nodes receive the *HopDistance* from beacon node who has the least hops between them. In the end of this step, unknown nodes compute the distance to the beacon nodes based hop-length and hops to the beacon nodes.

2.2 Introduction of Adaline

In this study, the neural network model was tested and optimized to obtain the best model configuration for the prediction of the roots of formula (3). An Adaline network typically comprises three types of neuron layers: an input layer, one or more hidden layers and an output layer each including one or several neurons. As shown in Figure 1, nodes from one layer are connected to all nodes in the following layer, but no lateral connections within any layer, nor feed-back connections are possible. Several input neuron are used, each representing an environmental variable. The output layer comprises one neuron. With the exception of the input neurons, which only connect one input value with its associated weight values, the net input for each neuron is the sum of all input values x_n , each multiplied by its weight w_{jn} , and a bias term z_j which may be considered as the weight from a supplementary input equalling one:

$$a_j = \sum w_{ji}x_i + z_j \tag{8}$$

The output value, y_j , can be calculated by feeding the net input into the transfer function of the neuron:

$$y_j = f(a_j) \tag{9}$$

Many transfer functions can be used. In this study, two types of sigmoid functions have been compared: the tangential and logarithmic sigmoid transfer function^[16].

In this paper, we use one model of neural networks i.e., Adaline network, which is selected among the main neural network architectures used in engineering. The basis of the model is neuron structure as shown in Figure 1, where r is the number of elements in input vector. Each input is weighted with an appropriate X . The sum of the weighted inputs and the bias, forms the input to the transfer function f . Neurons may use any differentiable transfer function f to generate their output. The transfer function of Adaline is shown in Figure 2.

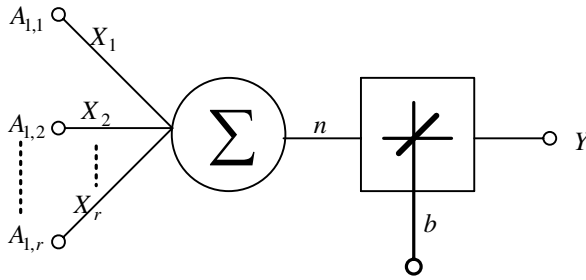


Fig. 1. An elementary neuron with r inputs

Artificial Neural Networks (ANNs) consist of a great number of processing elements (neurons), connected to each other. The strengths of the connections are called weights. A feedforward multilayer network is commonly used to model physical systems. It consists of a layer of input neurons, a layer of output neurons and one or more hidden layers.

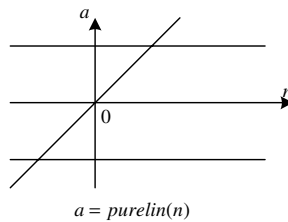


Fig. 2. Linear transfer function

Widrow and Hoff’s Adaline is one of the most effective, most well understood, and most widely used connectionist learning components (Widrow and Hoff, 1960). This paper introduces the Adaline model to estimate the roots of formula (3), because it is linear relation for formula (3).

The error function is defined as follows.

$$E(W, B) = \frac{1}{2}(T - A)^2 = \frac{1}{2}(T - WP)^2 \tag{10}$$

In order to minimize the error function (10), we use the *W-H* rules. At the same time, in order to optimize the network, we must exercise it, the process is as follows.

- 1) Calculating output vector $A = W \times P + B$ and error $E = T - A$ between expectation and average value;
- 2) Comparing between output error and expectation and minimizing the output error.
- 3) Introducing *W - H* rule to calculate new weight and error and return 1).

In order to use the Adaline network, the formula (3) is written as

$$\begin{cases} \min [sum(A(1)) \quad sum(A(2))]X - sum(b) \\ s.t. AX = b \end{cases} \tag{11}$$

We introduce the Adaline network to resolve the optimized roots of formula (11).

2.3 Measuring Hop-Count Using DV-Hop

Each of the anchor nodes launches the DV-Hop algorithm by initiating a broadcast containing its known location and a hop count of 0. All of the one-hop neighbors surrounding the anchor hear this broadcast, record the anchor’s position and a hop count of 1, and then perform another broadcast containing the anchor’s position and a hop count of 1. Every node that hears this broadcast and did not hear the previous broadcasts will record the anchor’s position and a hop count of 2 and then rebroadcast. This process continues until each anchor’s position and an associated hop count value have been spread to every node in the network. It is important that nodes receiving these broadcasts search for the smallest number of hops to each anchor. This ensures conformity with the model used to estimate the average distance of a hop, and it also greatly reduces network traffic. One model for estimating the average hop distance between nodes for the entire network is to simply use the maximum radio range of each node. This simplistic approach is sufficient to generate satisfactory position results, and saves on communication costs relative to more complicated models. The details are shown in Figure 3.

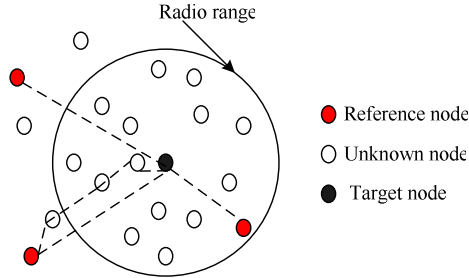


Fig. 3. Illustration of calculating hop-count

Once a node has received data regarding at least three(four) anchor nodes for a network existing in a two(three)-dimensional space, it is able to perform a ADV-Hop to estimate its location. If this node subsequently receives new data after already having performed a ADV-Hop, either a smaller hop count or a new anchor, the node simply performs another ADV-Hop to include the new data. This procedure is summarized in the following piece of pseudo code:

```

when a positioning packet is received,
  if new anchor or lower hop count then
    store (hop count + 1) for this anchor.
    compute estimated range to this anchor.
    broadcast new packet for this anchor.
    
```



```

else
  do nothing.
if number of anchors ≥ (dimension of space + 1) then
  ADV-Hop.
else
  do nothing.

```

In order to estimate the performance of algorithm, we define the location error ϵ as

$$\epsilon = \sqrt{(X_{est} - X_i)^2 + (Y_{est} - Y_i)^2} \tag{12}$$

where (X_i, Y_i) is actual location of receiver, (X_{est}, Y_{est}) is estimated location of the receiver using formula (11).

Furthermore, we define the location average error η as

$$\eta = \frac{\sum_i \sqrt{(X_{est}^i - X_i)^2 + (Y_{est}^i - Y_i)^2}}{K} \tag{13}$$

where (X_i, Y_i) is actual location of receiver i , (X_{est}^i, Y_{est}^i) is estimated location of the receiver i using formula (11) and K is number of unknown nodes.

3 Simulation Results

To validate our improved method, we consider an experiment region of square area of 50m×50m and sensor nodes are assumed to be randomly distributed in that area. The number of sensor nodes and the radio range of sensor nodes will be varied. We have implimented a number of experiments to cover a wide range of algorithm configurations including varying the ratio of anchor nodes, the number of unknown nodes, and the radio range.

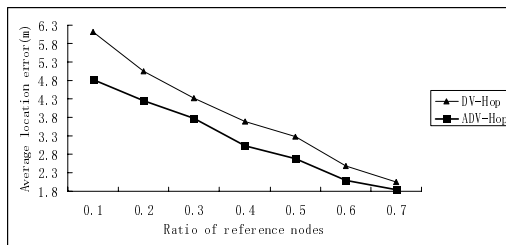


Fig. 4. Average location error versus ratio of reference nodes

Figure 4 shows the variation of the average localization errors as the reference ratios. In this experiment, the number of sensor nodes is fixed to 200. Suppose that the total estimation error is the summation of the Euclidean errors between true positions and estimated positions of all unknown nodes. Here the average localization error is defined as ratio between the total error and the number of the unknown nodes. Figure 4 is

obtained by averaging over 100 dependent network simulations. It can be seen from Figure 4 that the average localization error by the ADV-Hop algorithm is obviously less than the DV-Hop method in all considered conditions. For example, with 20 anchor nodes(20%), our ADV-Hop has an average error of about 4m, whereas the DV-Hop has an average error of about 5m.

The number of unknown nodes affects the ADV-Hop algorithm. In this experiment, the number of anchor nodes is fixed to 50. We can see from Figure 5 that the location error of these two algorithms is decreased with increasing the number of unknown nodes. This is because with the increase of unknown nodes, the node density in networks is increased, consequently the average number of neighbors is also increased. Thus, the network will be well connected and has a higher connectivity. This increases probability that there exist unknown nodes located on the line between anchor node i and j in each broadcast of hop count. Then the average hopdistance estimated by any pair of anchor nodes will be accurate and thus the estimated distance between the unknown node and the anchor node using average *HopDistance* will be closer to the true distance between the unknown node and the anchor node. So the location error of the algorithm is decreased with increasing the number of unknown nodes. Our ADV-Hop algorithm also achieves better performance than the DV-Hop in the scenario.

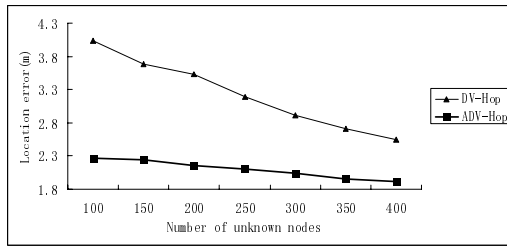


Fig. 5. Location error versus number of unknown nodes

Figure 6 shows the cumulative density function of location error. Here, the number of unknown nodes is fixed to 50 and the number of anchor nodes is fixed to 50. Over 96% of the nodes have less than a 3-meter error in our proposed algorithm, compared to 80% of nodes for the conventional DV-Hop algorithm.

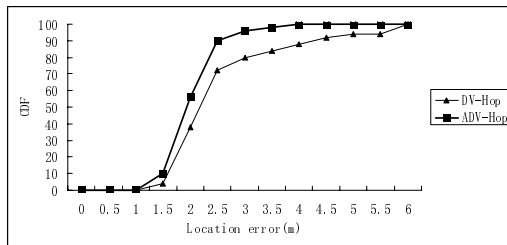


Fig. 6. Cumulative distribution functions of the estimation error

Figure 7 shows the location error of each approach for different radio ranges. Here, the number of unknown nodes is fixed 80 and the number of anchor nodes is fixed to 20. The location accuracy increases as the radio range increases. This increase in location accuracy is due to the fact that the estimated error between estimated distance using average *HopDistance* and true distance between the unknown node and the anchor node is increased. Considering that the connectivity of sensor nodes can be controlled by specifying its radio range, an increase to the radio range leads to an increase of network connectivity. Consequently, the number of neighboring anchor nodes per unknown node will also increase.

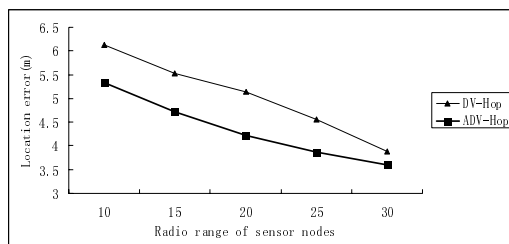


Fig. 7. Location error versus radio range

4 Conclusions

In this paper, we proposed an improved DV-Hop algorithm for locating the unknown nodes, called ADV-Hop. A simple estimation model is developed by introducing Adaline network to estimate the optimized roots of formula (3). The simulation results show that our proposed method can reduce the nodes average localization error significantly in different communication ranges and reference node ratios.

Acknowledgments

The author would like to thank the Science and Technology Research Project of Chongqing Municipal Education Commission of China under Grant No. 080526 and the Doctoral Research Fund of Chongqing University of Posts and Telecommunications (A2006-08). The author would also like to thank MATLAB software.

References

1. Chen, P.C.: A Nonlinear-of-sight Error Mitigation Algorithm in Location Estimation. In: Proc. IEEE Wireless Communication. Networking Conf., September 1999, pp. 316–320 (1999)
2. Spirito, M.A.: On the Accuracy of Cellular Mobile Station Location Estimation. IEEE Trans. Veh. Technol. 50, 674–685 (2001)

3. Reed, J.H., Krizman, K.J., Woerner, B.D., Rappaport, T.S.: An Overview of the Challenges and Progress in Meeting the E-911 Requirement for Location Service. *IEEE Communication Magazine*, 30–37 (1998)
4. Werb, J., Lanzl, C.: Designing a Positioning System for Finding Things and People Indoors. *IEEE Spectrum* 35, 71–78 (1998)
5. Ward, A., Jones, A., Hopper, A.: A new Location Technique for the Active Office. *IEEE Pers. Communication* 4, 42–47 (1997)
6. Torrieri, D.J.: Statistical Theory of Passive Location Systems. *IEEE Trans. On AES* 20(2), 183–198 (1984)
7. Rappaport, T.S.: *Wireless Communications: Principles and Practice*, pp. 50–143. Prentice-Hall, New Jersey (1996)
8. Girod, L., Estrin, D.: Robust Range Estimation using Acoustic and n Multimodal Sensing. In: *IEEE International Conference on Intelligent Robots and Systems*, vol. 3, pp. 1312–1320 (2001)
9. Cheng, X., Thaler, A., Xue, G., Chen, D.: TPS: a Time-based Positioning Scheme for Outdoor Wireless Sensor Networks. In: *IEEE INFOCOM 2004*, Hong Kong, China, pp. 2685–2696 (2004)
10. He, T., Huang, C., Blum, B.M., Stankovic, J.A., Abdelzaher, T.: Range-Free Localization Schemes for Large Scale Sensor Networks. In: *Proc. of the ACM MobiCom 2003*, San Diego, pp. 81–95 (2003)
11. Savvides, A., Han, C.C., Srivastava, M.: Dynamic Fine-grained Localization in Ad-hoc Networks of Sensors. In: *Proceeding of the 7 th ACM International Conference on Mobile Computing and Networking (MOBICOM)*, Rome, Italy, pp. 166–179 (2001)
12. Chen, J., Yao, K., Hudson, R.: Source Localization and Beamforming. *IEEE Signal Processing Magazine* 19, 30–39 (2002)
13. Čapkun, S., Hamdi, M., Hubaux, J.P.: Gps-Free Positioning in Mobile Ad Hoc Networks. In: *Proc. of Hawaii Int'l. Conf. System Sciences*, pp. 3481–3490 (2001)
14. Niculescu, D., Nath, B.: Ad Hoc Positioning System (APS). In: *Proc. of the IEEE GLOBECOM 2001*, San Antonio, pp. 2926–2931 (2001)
15. Doherty, L., Pister, K., Ghaoui, L.E.: Convex Position Estimation in Wireless Sensor Networks. In: *IEEE INFOCOM 2001*, Anchorage, AK (2001)
16. Cong, S.: *Object MATLAB Box Nerve Network Theory and Application*, pp. 31–40. Science and Technology of University of China Press, Hefei (1998)

Remote Estimation with Sensor Scheduling

Li Xiao, Zigang Sun, Desen Zhu, and Mianyun Chen

Key Laboratory of Ministry of Education for Image Processing and Intelligent Control
Department of Control science and Engineering, Huazhong University of Science and
Technology, Wuhan 430074, China
{Li Xiao, Zigang Sun, Desen Zhu, Mianyun Chen, wh_xl}@163.com

Abstract. A time-varying Kalman filter is proposed to solve the problem of remote estimation with sensor scheduling and measurement loss. The statistical properties of the estimation error are studied. The expectation of the estimation error covariance is proved to have upper and lower bounds. Convergence conditions and methods to calculate these bounds are also presented. The optimal sensor selection probability is found by using gradient search method. When the remote estimator schedules the transmission of sensors using optimal probability, the best estimation performance can be obtained. The validity of the proposed results are demonstrated by numerical examples.

Keywords: Kalman Filter, Sensors Scheduling, Remote Estimation.

1 Introduction

Recently, there is a growing interest in applying distributed estimation to wireless sensor networks, networked control systems, distributed control systems, et, al [1,2]. Typical advantages of using distributed estimation include relatively lower costs, inherent robustness and greater mobility. In distributed estimation, measurements from sensors are packed and transmitted over wireless channels, and for the limited spectrum, time varying channel gains and interferences, the wireless channels are not reliable. Packet loss is inevitable because of collisions and transmission errors. Estimation over packet loss networks has recently been studied in [3-7], where estimator was assumed to use one sensor with intermittent measurements.

In this paper, we extend the results in [4] to the case of using multi-sensors. In this case, sensors are scheduled to send measurements by a scheduler which adopting stochastic selection algorithm as in [2]. A remote estimator uses scheduled and intermittent measurements to estimate process states. When arrival probabilities of all sensors' measurement packets are known, the optimal estimation can be obtained if the scheduler uses the optimal sensors scheduling. The rest of this paper is organized as follows: in section 2, the implementation of the Kalman filter with sensor scheduling is proposed. The estimation error is stochastic due to the sensor scheduling and packet loss, and the bounds of estimation error covariance and convergence conditions are studied in section 3. A gradient search method which is used to find the optimal sensor selection probability is presented in section 4. And numerical examples are provided in section 5 to show the validity of the proposed approaches.

2 Implementation of Remote Estimation

We consider a discrete-time process with dynamics:

$$x(k+1) = Ax(k) + Bw(k), \tag{1}$$

where $x(k) \in \mathbf{R}^n$ is the process state, the initial state x_0 is Gaussian with mean \bar{x}_0 and covariance matrix P_0 . Process noise $w(k)$ is assumed to be white Gaussian noise with zero mean and covariance matrix Q .

The outputs of the process are measured by N sensors according to the equation:

$$y_i(k) = C_i x(k) + v_i(k) \quad (i = 1, \dots, N), \tag{2}$$

where $y_i(k) \in \mathbf{R}^m$ and $v_i(k) \in \mathbf{R}^p$ are measurement and measurement noise of the i th sensor respectively. The noises $v_i(k)$ are also assumed to be white Gaussian with zero mean and covariance matrix R_i . The schematic diagram for remote estimation with sensor scheduling is shown in figure1.

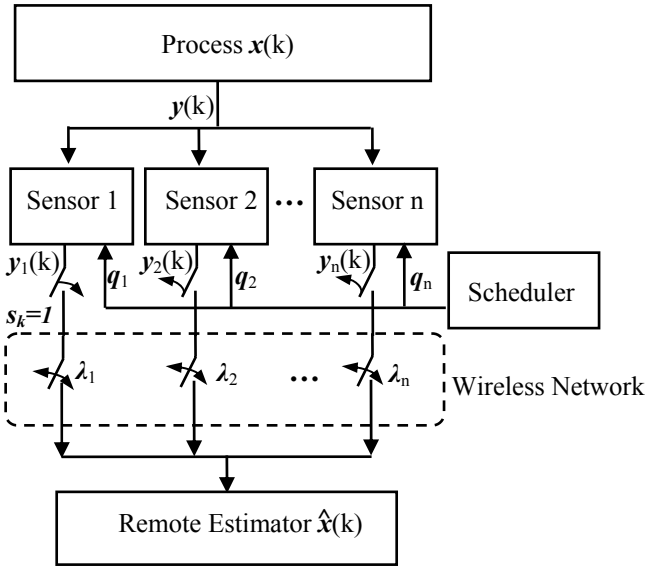


Fig. 1. Schematic diagram of remote estimation with sensor scheduling

The sensors are scheduled to send measurements by a scheduler using stochastic selection algorithm. At every time step, the i th sensor is chosen independently with probability q_i , and then transmits the measurement to remote estimator over wireless channels in one data packet. Sensor scheduling results in remote estimator switching between sensors. We use $\{s_k\}$ to denote sensor selection sequence where $s_k \in [1, N]$ satisfy $\Pr(s_k=i)=q_i$. Because of being transmitted over unreliable wireless channels, packet loss is inevitable. Binary random variable $\gamma_{i,k}$ indicates at time step k whether measurement from i th sensor is received correctly. We assume that $\gamma_{i,k}$ is i.i.d.

Bernouli process with $\Pr(\gamma_{i,k} = 1 | s_k = i) = \lambda_i$, where λ_i is the probability of measurements from i th sensor transmitting correctly. At different time step k and s i.e. $k \neq s$, $\gamma_{i,k}$ and $\gamma_{i,s}$ are independent. We also assume that arrivals of the measurement from all sensors are all random processes. The measurement noise of i th sensor is defined below [4].

$$P(v_{i,k} | \gamma_{i,k}) = \begin{cases} N(0, R_i), & \gamma_{i,k} = 1, \\ N(0, \sigma_i^2 I), & \gamma_{i,k} = 0. \end{cases} \quad (3)$$

At time step k , scheduler chooses i th sensor to transmit measurement, if packet reaches the remote estimator i.e. $\gamma_{i,k} = 1$, the covariance of measurement noise is R_i , otherwise $\gamma_{i,k} = 0$ and covariance is σ_i^2 . In practice, the absence of measurement corresponds to the limiting case of $\sigma_i \rightarrow \infty$.

The estimator proceeds similarly to the standard Kalman filter in [8] except that the correction step. In this step, if the measurement packet from selected sensor arrives, the estimator uses this measurement and the matrix of the selected sensors to correct the state estimation, if measurement packet is lost, estimator run in open loop. We use $P_{k+1,i} = P_{k+1|k,i}$ to denote the error covariance at time step $k+1$ under condition of using measurement from i th sensor at k time step. The equation of $P_{k+1,i}$ is

$$P_{k+1,i} = AP_k A^T + BQB^T - \gamma_{i,k} AP_k C_i^T (C_i P_k C_i^T + R_i)^{-1} C_i P_k A^T, \quad (4)$$

which can be expressed as

$$\begin{cases} P_{k+1,i} = f_{C_i}(P_k) = AP_k A^T + BQB^T - AP_k C_i^T (C_i P_k C_i^T + R_i)^{-1} C_i P_k A^T & \gamma_{i,k} = 1, \\ P_{k+1,i} = f_{i,loss}(P_k) = AP_k A^T + BQB^T & \gamma_{i,k} = 0. \end{cases} \quad (5)$$

We use a simplified notation $P_k = P_{k|k-1}$, and since error covariance P_k depends on sensor selection sequence $\{s_k\}$ and packet arrival sequence $\{\gamma_{i,k}\} (i=1, \dots, N)$ which are stochastic processes, P_k is stochastic variable. So we study its expected value $E[P_k]$ and try to evaluate it in the limiting case of $k \rightarrow \infty$.

$$\bar{P}_k = E[P_k] = \sum_{i=1}^N q_i E[P_{k,i}], \quad (6)$$

In general, it's intractable to evaluate the expected value $E[P_k]$ explicitly, but we can find its upper and lower bounds.

3 Bounds and Convergence Conditions

To study the convergence properties of $E[P_k]$ under any initial conditions, we extend the results in [4] to the case of estimation with sensor scheduling and measurement dropping. The theorems below give the upper and lower bounds to $E[P_k]$.

Theorem 1 (upper bound). N sensors are used to measure the process output, and only one sensor is randomly chosen at every time step. If the i th sensor is chosen independently at each time step with probability q_i , and the arrival probability of the

measurements from i th sensors is λ_i , the upper bound of the expected error covariance is given by the recursion

$$\bar{P}_{k+1} = A\bar{P}_kA^T + BQB^T - \sum_{i=1}^N q_i \lambda_i \left[A\bar{P}_kC_i^T (R_i + C_i\bar{P}_kC_i^T)^{-1} C_i\bar{P}_kA^T \right], \tag{7}$$

where $\bar{P}_0 = P_0$.

Proof. The arrival probability of i th sensor is λ_i , so estimator uses the measurements of the i th sensor with probability $q_i\lambda_i$, the loss probability of the measurements of i th sensor is $q_i(1-\lambda_i)$. As selection of sensors is independent, if take the expectation to (4) with respect to the probability distribution of C_i , it results

$$E[P_{k+1}] = \sum_{i=1}^N q_i \left[\lambda_i E[f_{C_i}(P_k)] + (1-\lambda_i) E[f_{i,loss}(P_k)] \right].$$

Using Jensen’s inequality [9], we can obtain

$$\begin{aligned} \bar{P}_{k+1} &= E[P_{k+1}] = \sum_{i=1}^N q_i \left[\lambda_i E[f_{C_i}(P_k)] + (1-\lambda_i) E[f_{i,loss}(P_k)] \right] \\ &\leq \sum_{i=1}^N q_i \lambda_i f_{C_i}(E[P_k]) + \sum_{i=1}^N q_i (1-\lambda_i) f_{i,loss}(E[P_k]) \\ &= A\bar{P}_kA^T + BQB^T - \sum_{i=1}^N q_i \lambda_i \left[A\bar{P}_kC_i^T (R_i + C_i\bar{P}_kC_i^T)^{-1} C_i\bar{P}_kA^T \right], \end{aligned}$$

thus concluding the proof.

Theorem 2 (convergence condition of upper bound). If there exist matrices $\tilde{K}_1, \tilde{K}_2, \dots, \tilde{K}_N$ and a positive definite matrix \bar{P} such that

$$\bar{P} > \sum_{i=1}^N q_i \left[(1-\lambda_i)(AXA^T + BQB^T) + \lambda_i(F_iXF_i^T + V_i) \right], \tag{8}$$

where $F_i = A + \tilde{K}_iC_i$, $V_i = \tilde{K}_iR_i\tilde{K}_i^T + BQB^T$. Then the iteration in (7) converges for all initial conditions $P_0 \geq 0$, and the limit \bar{P} is the unique positive semi-definite solution to the equation

$$X = AXA^T + BQB^T - \sum_{i=1}^N q_i \lambda_i A \left[XC_i^T (R_i + C_iXC_i^T)^{-1} C_iX \right] A^T. \tag{9}$$

Proof. Define Modified Algebraic Riccati Equation (MARE) and operator function to every sensor:

$$g_i(X) = AXA^T + BQB^T - \lambda_i AXC_i^T (C_iXC_i^T + R_i)^{-1} C_iXA^T, \tag{10}$$

$$\phi_i(\tilde{K}_i, X) = (1-\lambda_i)(AXA^T + BQB^T) + \lambda_i \left[F_iXF_i^T + V_i \right] \tag{11}$$

where $F_i = A + \tilde{K}_iC_i$, $V_i = \tilde{K}_iR_i\tilde{K}_i^T + BQB^T$. Because all sensors are scheduled by stochastic algorithm, the MARE and operator function of the remote estimator become

$$\begin{aligned}\tilde{\phi}(\tilde{K}_1, \dots, \tilde{K}_N, X) &= \sum_{i=1}^N q_i \phi_i(\tilde{K}_i, X) \\ &= \sum_{i=1}^N q_i \left[(1 - \lambda_i)(AXA^T + BQB^T) + \lambda_i(F_i X F_i^T + V) \right], \\ \tilde{g}(X) &= \sum_{i=1}^N q_i g_i(X) = AXA^T + BQB^T - \sum_{i=1}^N q_i \lambda_i AXC_i^T (C_i X C_i^T + R_i)^{-1} C_i X A^T.\end{aligned}$$

Define $L(Y) = \sum_{i=1}^N q_i \left[(1 - \lambda_i)AXA^T + \lambda_i F_i Y F_i^T \right]$, and follow the arguments given in proof of theorem 1 in [4], we can prove the theorem. \square

When only one sensor is adopted to measure the process output, the upper bound in theorem 1 is same as the bound in theorem 1 in [4]. When there are no measurement losses to every sensor, i.e. $\lambda_i=1(i=1, \dots, N)$, theorem 2 is similar to the theorem 3 in [2]. If arrival probabilities of all sensors are known, we can calculate the optimal sensor selection probability by using gradient search algorithm.

Theorem 3 (lower bound). Let $\tilde{A} = \sqrt{1 - \sum_{i=1}^N q_i \lambda_i} A$, there exists a matrix sequence $\{S_k \mid S_0 = 0, S_k = \tilde{A}S_{k-1}\tilde{A}^T + BQB^T\}$ which satisfies $S_k \leq \bar{P}_k$. If (\tilde{A}, C_i) is stable and $(\tilde{A}, BQ^{\frac{1}{2}})$ is controllable, the matrix sequence S_k converges to matrix \bar{S} as $k \rightarrow \infty$.

Proof. Define a Lyapunov operator $m(X) = \tilde{A}X\tilde{A}^T + BQB^T$, then $S_0 = 0 \leq BQB^T = S_1$. From lemma 2 in [4], $m(\cdot)$ is monotonically increasing, i.e. $S_{k+1} > S_k$ to all steps. If (\tilde{A}, C_i) is stable, then as $k \rightarrow \infty$, $\{S_k\}$ is convergence. The limit \bar{S} is the solution to the Lyapunov equation $\bar{S} = \tilde{A}\bar{S}\tilde{A}^T + BQB^T$. When $(\tilde{A}, BQ^{\frac{1}{2}})$ is controllable, the solution is positive.

Let $\tilde{K}_i = -A\tilde{P}C_i^T(C\tilde{P}C_i^T + R_i)^{-1}$, from lemma 1 in [4], the function $g_i(X)$ (10) and $\phi_i(\tilde{K}_i, X)$ (11) satisfy

$$\begin{aligned}g_i(X) &= \phi_i(\tilde{K}_i, X) \\ &= (1 - \lambda_i)(AXA^T + BQB^T) + \lambda_i \left[(A + \tilde{K}_i C_i)X(A + \tilde{K}_i C_i) + BQB^T + \tilde{K}_i R_i \tilde{K}_i^T \right] \\ &\geq (1 - \lambda_i)(AXA^T + BQB^T) + \lambda_i BQB^T.\end{aligned}$$

Take the expectation in (10) with respect to the probability distribution of C_i , we can get

$$\begin{aligned}\bar{P}_k &= \tilde{g}(X) = \sum_{i=1}^N q_i g_i(X) \\ &\geq \sum_{i=1}^N q_i [(1 - \lambda_i)(AXA^T + BQB^T) + \lambda_i BQB^T] \\ &= \sum_{i=1}^N q_i (1 - \lambda_i)AXA^T + BQB^T = S_k,\end{aligned}$$

thus concluding the proof.

The theorem 1 to theorem 3 give the bounds of $E[P_k]$, and can be used to evaluate the performance of the remote estimator. Next, we use the upper bound to solve problem of how to obtain optimal estimation.

4 Optimal Sensor Scheduling

The upper bound of $E[P_k]$ converges to a stead matrix X under conditions in theorem 2. When scheduler alters the sensor selection probability, the stead value also changes. The optimal sensor scheduling means to find a probability distribution so that the estimation error is smallest. In this paper $trace(X)$ is used as an approximation to the expected error covariance. Then the problem of optimal scheduling is

$$\min_{\mathbf{q}} trace(X) \text{ s.t. (9).} \tag{12}$$

The cost function is $trace(\tilde{g}_{\mathbf{q}}(X))$. Gradient search method can be used to find the optimal probability vector $\mathbf{q}=[q_1, \dots, q_N]^T$. Firstly, we define the function below

$$\tilde{g}_{\mathbf{q}}(X) = AXA^T + BQB^T - \sum_{i=1}^N \lambda_i q_i AX C_i^T (C_i X C_i^T + R_i)^{-1} C_i X A^T . \tag{13}$$

The algorithm proceeds as follows:

1. Initialization:

At step $k=1$, select an arbitrary valid probability vector $\mathbf{q}(1)$ s.t. $\sum_{i=1}^N q_i = 1, q_i > 0$, then calculate the positive semi-definite matrix $X(1)$ which satisfying $X(1) = \tilde{g}_{\mathbf{q}(1)}(X(1))$.

2. Searching optimal probability:

At every step k , calculate the valid probability vector \mathbf{q}_{\min} which can minimizes $\tilde{g}_{\mathbf{q}}(X(k))$. Probability vector \mathbf{q}_{\min} is obtained by solving an optimization problem below

$$\begin{aligned} & \arg \min_{\mathbf{q}} trace(AXA^T + BQB^T - \sum_{i=1}^N \lambda_i q_i AX C_i^T (C_i X C_i^T + R_i)^{-1} C_i X A^T), \\ & \sum_{i=1}^N q_i = 1, q_i > 0, \end{aligned}$$

where X is the given positive semi-definite matrix. This is a linear program and can be solved efficiently.

3. Update:

- Calculate $\bar{\mathbf{q}} = \mathbf{q}(k) + \delta(\mathbf{q}_{\min} - \mathbf{q}(k))$, and step size parameter $\delta \in (0, 1]$ is determined in advance.
- Project $\bar{\mathbf{q}}$ to $\mathbf{q}(k+1)$ which is in the set of valid probability vectors. This step is needed because the probability vector $\bar{\mathbf{q}}$ may not be a valid probability vector.

- Calculate $X(k+1) = \tilde{g}_{\mathbf{q}(k+1)}(X(k))$.
4. If $\mathbf{q}(k) = \mathbf{q}(k+1)$ then break, else $k:=k+1$, and then go to setp 2.

Vector $\mathbf{q}(k+1)$ is the optimal probability vector which can minimize the cost function $trace(\tilde{g}_{\mathbf{q}}(X))$. Convergence of the algorithm can be proved similarly to the one in [9],[10]. When sensors are scheduled under this probability vector, the performance of the remote estimator is optimal.

5 Numerical Examples

We assume a target moving in 2-D space according to standard constant acceleration model [11]. The positions of the target in the two planes are denoted by p_x and p_y , the velocities of the target are denoted by v_x and v_y . The discrete-time dynamics is

$$x_{k+1} = diag[A_x \ A_y]x_k + diag[B_x \ B_y]w_k,$$

where $A_x = \begin{bmatrix} 1 & h \\ 0 & 1 \end{bmatrix}, B_x = \begin{bmatrix} h^2/2 \\ h \end{bmatrix}, \mathbf{x}=[p_x \ v_x \ p_y \ v_y]^T$ is the system state, h is discretization step size, $w_k = [w_x \ w_y]^T$ is a discrete-time white noise sequence, w_x and w_y correspond to noisy accelerations along x and y axes respectively. It is assumed that $h=0.2s$, and the covariance matrix of process noise is $Q = \begin{bmatrix} 1 & 0.25 \\ 0.25 & 1 \end{bmatrix}$.

The measurement equations are

$$y_i(k) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} x(k) + v_i(k) \quad i = 1, 2.$$

Measurement noises $v_i(k)s$ are assumed to be white, zero mean and Gaussian, and are also independent from each other. The considered values of the sensor noise covariance are $R_1 = diag(2.4, 0.4)$ and $R_2 = diag(0.7, 1.4)$.

In example 1, it is assumed that sensor selection vector is $\mathbf{q} = [0.5 \ 0.5]^T$, the arrival probability of packers from sensor 1 and sensor 2 are 0.9 and 0.8 respectively. Estimations of the process state are illustrated in figure 2. The performance of the Kalman filter with sensor scheduling is very well.

Then theorem 1 and 2 are used to analysis how sensor selection and arrival rate affect error covariance. The arrival rates of the sensors are assumed to be const. Figure 3 shows the trace of upper bound in two cases that $\lambda_1=\lambda_2=1$ and $\lambda_1=0.9, \lambda_2=0.8$. It's obvious that there are local minimums in both cases. So if arrival rates are known, we can use gradient search algorithm to obtain optimal sensor selection probability. Figure 3 also shows that when the probability of measurement loss increases, the performance of the remote estimator degrades.

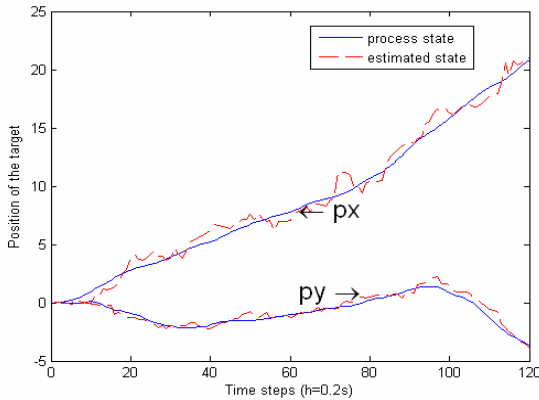


Fig. 2. Process state and estimated state with sensor scheduling ($q_1 = q_2 = 0.5$) and packets dropping ($\lambda_1 = 0.9, \lambda_2 = 0.8$)

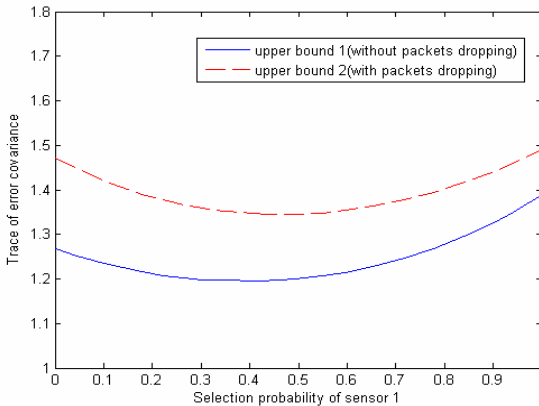


Fig. 3. Upper bound of estimation error with packets dropping ($\lambda_1 = 0.9, \lambda_2 = 0.8$) and without packets dropping ($\lambda_1 = \lambda_2 = 1$) respectively

Under two network conditions in the example 1, the optimal sensor selection probabilities are obtained by solve the optimal problem (12), and they are $\mathbf{q}_{opt1}=[0.37 \ 0.63]^T$ and $\mathbf{q}_{opt2}=[0.46 \ 0.54]^T$ respectively. Figure 4 and figure 5 plot the trace of time sequences of error covariance when the scheduler only choose sensor 1, only choose sensor 2 and use optimal sensor selection probability, and also plot the trace of the expectation of error covariance which is obtained according recursion (7) with optimal scheduling. When there are no packets dropping (in figure 4), the error covariance converges quickly, the trace of optimal time sequence fluctuates along the upper bound of the expectation of estimation error. In figure 5, estimation performance degrades as packets dropping, especially at time steps when the measurement loss. In two figures, it shows that the estimation performance can be improved obviously when scheduling sensors with optimal sensor selection probability.

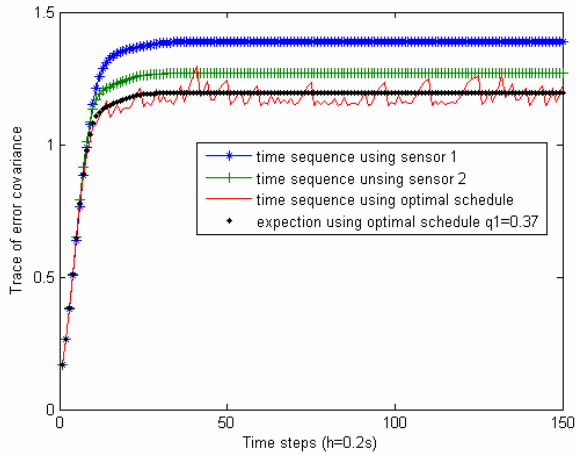


Fig. 4. Traces of error covariance when using different sensor selection probability and without packets dropping ($\lambda_1 = \lambda_2 = 1$)

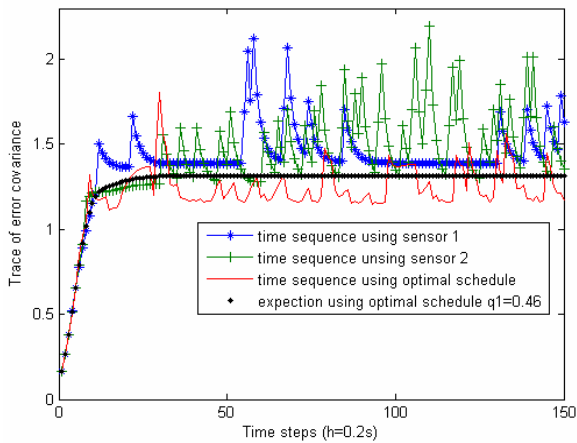


Fig. 5. Traces of error covariance when using different sensor selection probability and with packets dropping ($\lambda_1 = 0.9, \lambda_2 = 0.8$)

6 Conclusion

In this paper, remote estimation with sensor scheduling and measurement loss is presented. It shows that the exception of error covariance is bounded, and packet dropping will degrade the performance of the estimator. Gradient search algorithm is used to find the optimal sensor selection probability, and if sensors is scheduled according the optimal probability, the performance of the estimator is optimal.

Acknowledgments. The work was supported by the National Natural Science Foundation of China under the grant No. 60802002, the National Research Foundation for the Doctoral Program of Higher Education of China under Grant No. 20020487023.

References

1. Alriksson, P., Rantzer, A.: Sub-optimal Sensor Scheduling with Error Bounds. In: Proceedings of the 16th IFAC world congress, Prague (2005)
2. Gupta, V., Chung, T.H., Hassibi, B., Murray, R.M.: On a Stochastic Sensor Selection Algorithm with Applications in Sensor Scheduling and Sensor Coverage. *Automatica* 2, 251–260 (2006)
3. Smith, S., Seiler, P.: Estimation with Lossy Measurements: Jump Estimators for Jump Systems. *IEEE Trans. on Automatic Control* 48, 2163–2171 (2003)
4. Sinopoli, B., Schenato, L., et al.: Kalman Filtering with Intermittent Observations. *IEEE Transactions on Automatic Control* 49, 1453–1464 (2004)
5. Liu, X., Goldsmith, A.: Kalman Filtering with Partial Observation Losses. In: 43rd IEEE Conference on Decision and Control, Atlantis, Paradise Island, Bahamas, pp. 4180–4186 (2004)
6. Jin, Z., Gupta, V., Hassibi, B., Murray, R.M.: State Estimation Utilizing Multiple Description Coding over Lossy Networks. In: Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference 2005, Seville, Spain, pp. 872–878 (2005)
7. Hung, M., Dey, S.: Stability of Kalman Filtering with Markovian Packet Losses. *Automatica* 43, 598–607 (2007)
8. Astrom, K.J., Wittenmark, B.: *Computer Controlled Systems: Theory and Design*, 3rd edn. Prentice Hall Inc., Englewood Cliffs (1997)
9. Bertsekas, D.P., Nedic, A., Ozdaglar, A.E.: *Convex Analysis and Optimization*. Athena Scientific (2003)
10. Gupta, V.: *Distributed Estimation and Control in Networked Systems*. Ph.D. dissertation, California Institute of Technology (2006)
11. Li, X.R., Jilkov, V.P.: Survey of Maneuvering Target Tracking Part I: Dynamic Models. *IEEE Transactions on aerospace and Electronic Systems* 39(4), 1333–1363 (2003)

An Improved Margin Adaptive Subcarrier Allocation with Fairness for Multiuser OFDMA System

Tan Li^{1,2}, Gang Su¹, Guangxi Zhu¹, Jun Jiang¹, and Hui Zhang²

¹ Department of Electronics and Information Engineering,
Huazhong University of Science and Technology, Wuhan 430074, China

² State Key Laboratory of ISN, Xidian University, Xi'an, 710071 China
fanqielee@gmail.com, {gsu, gxzhu}@mail.hust.edu.cn,
little.fire.rock@hotmail.com, hzhang@xidian.edu.cn

Abstract. In this paper, according to the margin adaptive (MA) algorithm with fixed transmission rate, we propose a suboptimal MA Greedy algorithm with demand function. In this algorithm, the allocation process is divided into two separate steps, resource calculation and subcarrier allocation. With the employment of overload inhibition process, diversity scheduling method and demand function, we improve the traditional MA algorithms to reduce the computing complexity with subscribers' fairness and Quality of Service (QoS) guaranteed. Utilizing the simulation with real-time services, we have proved that compared to static allocation algorithms, our MA Greedy algorithm could enhance the system performance by 5-6dB with low computing complexity, which is very close to the optimal MA algorithm.

Keywords: Multiuser OFDMA systems, MA, Demand function, Fairness.

1 Introduction

Due to the ability to enhance the spectrum efficiency of OFDM systems, link adaptive (LA) transmission technology is considered as one of the key techniques of future wireless mobile communication systems. Recent researches[1,2] indicate that the combination of LA technology and OFDM can largely improve the performance of OFDM systems.

In multiuser OFDMA systems, the optimizing model of resource allocation would be quite complicated when both the QoS and the fairness constraint of subscribers are taking into consideration. High computing complexity is usually necessary for the optimal resource allocation schemes. However, this is very unfavorable to practical mobile communication systems because of the time-variant characteristics of wireless channels and the computing ability of communication equipments. Therefore, the suboptimal allocation algorithms with low computing complexity will be more promising for actual communication systems.

Lots of researches have already been carried out for suboptimal Greedy algorithm. Based on the Wong algorithm in [3, 4] has proposed a subcarrier allocation algorithm, the Wong2 algorithm, which could be applied to real-time systems. Although its performance is close to the optimal algorithm [5], its computing complexity is still considerable.

Reference [6] has provided another margin adaptive (MA) resource allocation algorithm. With its treatment for conflicting subcarriers, it could reduce the computing complexity effectively compared to the Wong algorithm. However, as it does not manage to control the transmitting power of subcarriers, its performance is impaired. Reference [7] has proved that it is optimal to allocate one subcarrier to just one subscriber for the consideration of transmission rate. Besides, the computing complexity is reduced successfully according to the expectation of the authors.

Consequently, we propose a suboptimal MA Greedy algorithm, which could increase the system capacity as much as possible while guaranteeing the proportional fairness of subscribers in OFDMA systems. The simulation with actual services model in this paper shows that our MA Greedy algorithm could enhance the system performance by 5-6dB with low computing complexity, which is very close to the optimal MA algorithm.

2 System Model and Objective

In the multiuser OFDMA system with N subcarriers and K subscribers [3] (shown in Fig. 1), it is assumed that one subcarrier is allocated to only one subscriber [7]. In this model, it is assumed that the power spectral density of additive Gaussian white noise

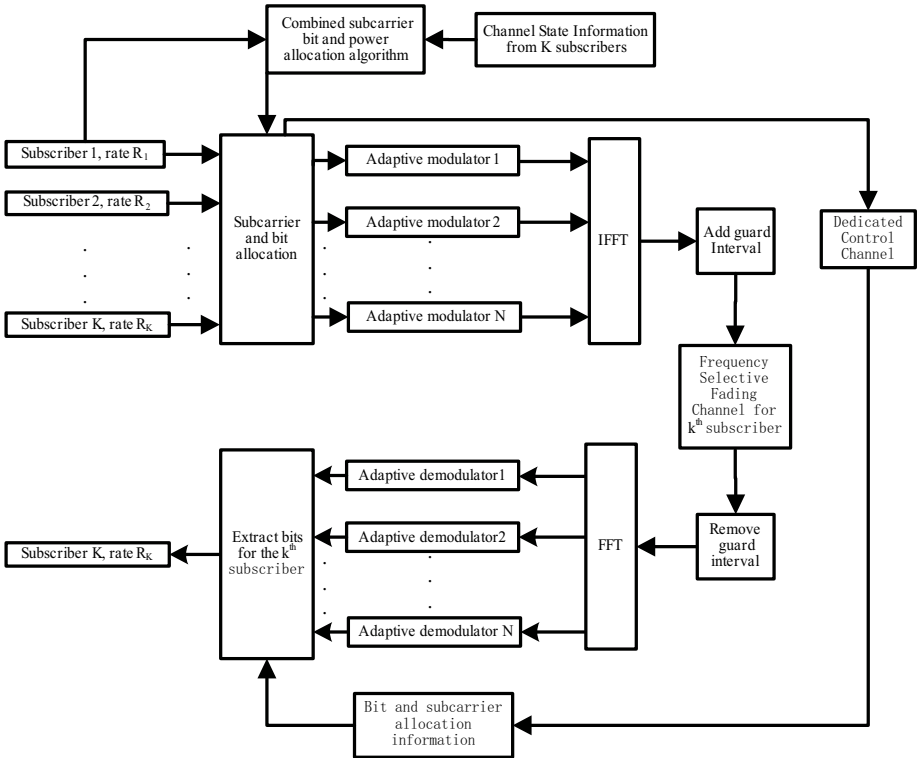


Fig. 1. The block diagram of multiuser OFDM system with subcarrier, bit and power allocation

is N_0 (w/Hz). R_k denotes the throughput of the k^{th} subscriber. H_{nk} is the transmission gain of the k^{th} subscriber on the n^{th} subcarrier which can be obtained from the feedback of the receivers. $f_k(c)$ is the required receiving power for the k^{th} subscriber to obtain c information bits reliably when the channel gain is equal to 1. P_T denotes the overall transmitting power of the systems. Meanwhile, it is assumed that the value of b_{nk} is taken from the set $\{0, 1, 2, \dots, M\}$, in which M is the maximal number of bits which could be allocated to one subcarrier with the highest modulation order and the maximal transmitting rate.

The objective of MA model is to minimize overall transmitting power with the constraints of limited bit error rate (BER) and the fixed transmission rate of subscribers. Therefore, its mathematical model can be presented as follows:

$$P_T^* = \min \sum_{n=1}^N \sum_{k=1}^K \frac{f_k(b_{nk})}{H_{nk}^2} \quad (1)$$

and the minimization is subjected to following constrains

$$\text{C1: } R_k = \sum_{n=1}^N b_{nk} \quad \text{for all } k \in \{1, \dots, K\} \quad (2)$$

$$\text{C2: For all } n \in \{1, \dots, N\},$$

$$\text{if } b_{nk'} \neq 0, b_{nk} = 0 \quad \text{for all } k \neq k' \quad (3)$$

3 The Suboptimal MA Greedy Algorithm Based on Demand Function

As shown in Figure 2, our suboptimal Greedy MA algorithm based on demand function is divided into two steps. With multiuser diversity scheduling [8], the first step is implemented in the resource calculation module, in which the modified BABS (Bandwidth Assignment Based on SNR) algorithm is employed. Its main functions are the evaluation of subscribers' demand and the quantity calculation of subcarriers which should be allocated to each subscriber. The second step is actualized in the subcarrier allocation module, in which the ACG (Amplitude Craving Greedy) algorithm based on

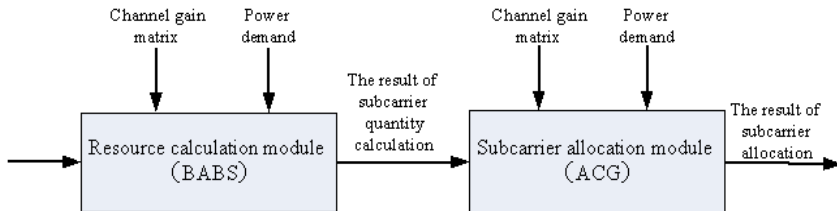


Fig. 2. The block diagram of the suboptimal MA Greedy algorithm based on demand function

demand function is utilized. In this step, each individual subcarrier is allocated according to channel status and the calculation result of the first step.

3.1 The Resource Calculation Module

In the resource calculation module, the total number of subcarriers which should be allocated to each subscriber is determined by channel status and the transmitting rate of the subscriber. Due to the feature of voice or multimedia services, that the performance of these services will largely decrease when their rates cannot reach the specified values, we would discard the subscribers who occupy a large number of subcarriers to guarantee the rate demand of most subscribers. Therefore, with the involvement of overload inhibition process, the modified BABS algorithm in this paper is more reasonable than the traditional BABS algorithm [9]. This modified algorithm is divided into two steps. In the first step, we allocate the minimal amount of subcarriers to each subscriber which may satisfy its rate demand. Then, the residual subcarriers will be allocated to the subscriber with lest power increment in the second step. The overload inhibition process is realized in the first step as follows:

1) Initialization

It is assumed that p denotes the total available subcarriers in the system, and it is initialized to N . n_k is the amount of subcarriers which should be allocated to the k^{th} subscriber, and it is initialized to 0.

2) Calculation

For all K subscribers,

$$n_k = \lceil R_k / M \rceil \quad (4)$$

$$p = p - n_k \quad (5)$$

3) Overload Inhibition

Repeat following steps until $p \geq 0$.

- a) The values of n_k are sorted in descending order in the set $\{n_k\}$, and $ind(s)$ denotes the subscriber number corresponding to the s^{th} element of this set.
- b) The subscriber who needs the largest number of subcarriers is discarded.

$$n_{ind(i)} = 0 \quad (6)$$

$$p = p + n_{ind(i)} \quad (7)$$

- c) The first element of $\{n_k\}$ is discarded, and the set $\{n_k\}$ is updated.

Then, in the second step, the residual subcarriers are allocated according to the power minimization principle. Herein, the discarded subscribers in the first step are not involved in this allocation process. At last, the amount of subcarriers which should be allocated to each subscriber is determined.

3.2 The Subcarrier Allocation Module

In the subcarrier allocation module, each specific subcarrier is allocated to a specific subscriber according to the channel status and the allocation result of the resource calculation module. To achieve a reasonable allocation of the subcarriers, we modify the allocation function of the traditional ACG algorithm with proposing the demand function of the k^{th} subscriber on the n^{th} subcarrier, as shown in (8).

$$fa\ var(k, n) = \frac{|H_{nk}|^2}{\sum_n |H_{nk}|^2} \quad (8)$$

In other words, utilizing $fa\ var(k, n)$ to indicate the demand degree of the k^{th} subscriber for the n^{th} subcarrier, and subcarriers are allocated to the subscriber who needs them most. The detailed steps are as follows:

1) Initialization

It is assumed that $al(n)$ denotes the number of the subscriber who obtains the n^{th} subcarrier, and it is initialized to -1. m_k is the total number of subcarriers which have already been allocated to the k^{th} subscriber, and it is initialized to 0.

2) Allocation

For each subcarrier, $fa\ var(k, n)$ is calculated for all K subscribers, and these K values are sorted in descending order in the set $\{fa\ var(k, n)\}$. In this set, we search from the first element to find out the first subscriber \bar{k} whose $m_{\bar{k}}$ is less than its $n_{\bar{k}}$, and then

$$al(n) \leftarrow \bar{k} \quad (9)$$

$$m_{\bar{k}} \leftarrow m_{\bar{k}} + 1 \quad (10)$$

Eventually, all the N subcarriers are allocated to K subscribers. Then, the Greedy algorithm for single subscriber is employed to achieve the optimal allocation results of subcarriers, bits and power.

4 Simulation and Performance Comparison

In this simulation for OFDMA systems, the frequency selective Rayleigh fading channel with 3 independent paths is employed. The power attenuation of each path can

Table 1. Service models

Service Type	Fixed Rate (Yes/No)	Base Rate (kb/s)	Min Rate (kb/s)	Max Rate (kb/s)	The Range of δ
Voice	Yes(1)	16	16*0.9	16*1.1	[0,0.3]
Video	Yes(1)	64	64*0.9	64*1.1	[0.3,0.6]
Data	No(0)	30	6	30*10	[0.6,1]

be described as the negative index distribution. K independent channels are contained in each group. At most 128 subcarriers are utilized in this system. It is assumed that the power spectral density of noise is 1, and the parameters of the service models are shown in Table 1.

Herein, we assume that there are 30% subscribers using voice service, and 30% using video service, 40% using data service. Therefore, subscribers choose their service model via a random variable δ . If $\delta \in [0,0.3)$, the voice service is selected. If $\delta \in [0.3,0.6)$, the video service is selected. If $\delta \in [0.6,1]$, the video service is selected (shown in Table 1). In the simulation of MA algorithm, voice and video services are utilized.

In order to analyze the performance of our proposed MA algorithm, the OFDM-TDMA and OFDM-BFDMA [10] static allocation algorithms are employed in this simulation.

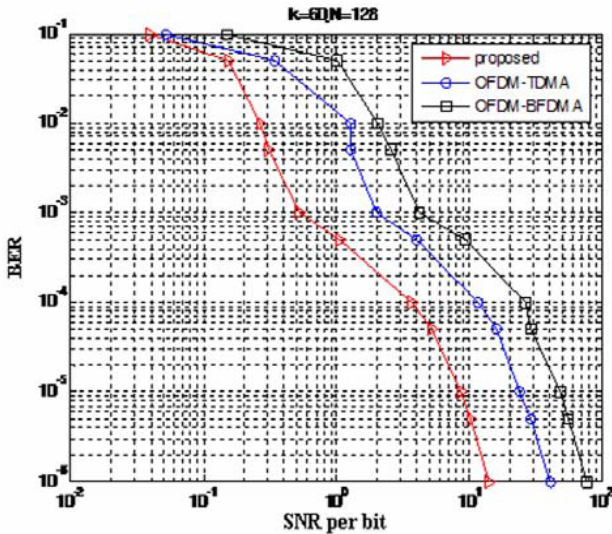


Fig. 3. BER versus the SNR per bit of different algorithms

Figure 3 shows BER versus the requisite SNR per bit of different algorithms. Herein, SNR per bit is defined as the transmitting power per bit divided by the power spectrum density of noise, in which the transmitting power per bit is the overall transmitting power, including the transmitting power of all the subcarriers and subscribers, divided by the amount of bits in an OFDM symbol. As shown in Figure 3, the performance of the proposed MA algorithm is 5-6 dB higher than that of the static allocation algorithms. Besides, as shown in the curve of the proposed algorithm, the BER decreases even faster with the increment of SNR per bit, because our dynamic MA allocation algorithm can utilize the mutative channel status information (CSI) well. However, the allocation result of the static algorithms is specified before the allocation process, so it is independent with BER [10].

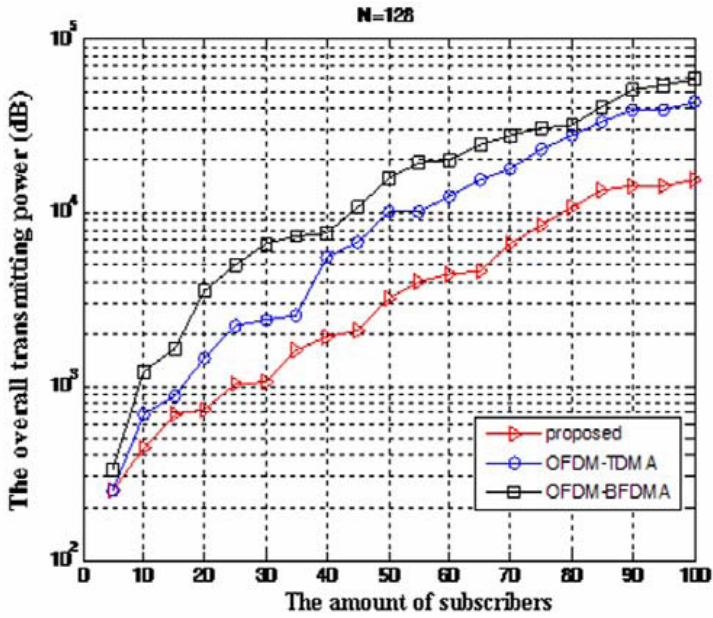


Fig. 4. The overall transmitting power versus the amount of subscribers

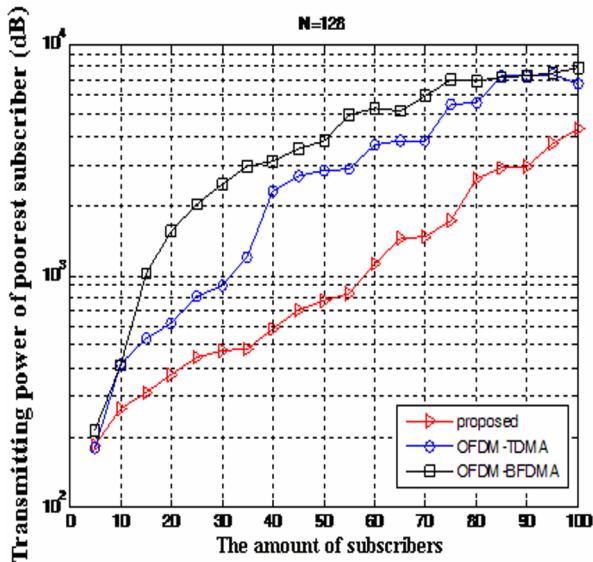


Fig. 5. The transmitting power of the poorest subscriber versus the amount of subscribers

In order to observe the influence of the amount of subscribers on the system performance, two figures are given by this simulation. Figure 4 shows the overall transmitting power versus the amount of subscribers, and Figure 5 shows the transmitting power

of the subscriber who is in the poorest channel status. As shown in these figures, the overall transmitting power of the proposed MA algorithm is much lower than that of the static algorithms, because our dynamic MA algorithm could utilize the diversity of subscribers well, and each subscriber could obtain the most suitable subcarriers. However, as the static algorithms pre-specify the allocation result, more transmitting power is needed to satisfy the BER constraint compared to the proposed MA algorithm, especially when FDMA is employed. As we know, in frequency selective fading environments, the correlation between channels is very high, and it is possible that all the subcarriers allocated to a subscriber are in deep fading status. Therefore, the transmitting power is very high under this condition when the FDMA static algorithm is employed.

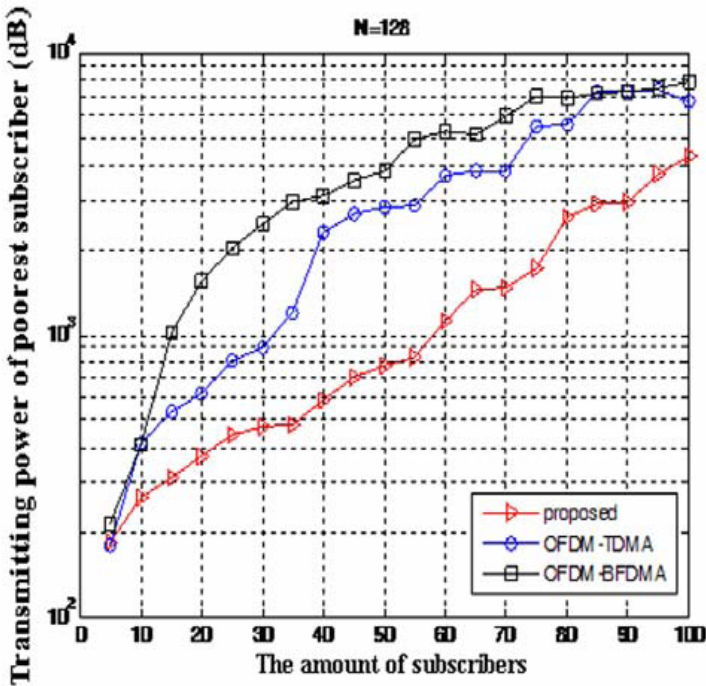


Fig. 6. The overall transmitting power versus the amount of subcarriers

Figure 6 shows the overall transmitting power versus the amount of subcarriers. Again, we can see that the performance of the proposed algorithm is 5-6 dB higher than that of the static algorithms. Moreover, as shown in the curve of the proposed algorithm, the overall transmitting power decreases even faster with the increment of subcarrier quantity, because our dynamic allocation algorithm can utilize the mutative channel status information (CSI) well, and the subscribers can choose more suitable subcarriers when the subcarrier quantity increases. However, the allocation result of the static algorithm is specified before the allocation process, so the proposed algorithm can achieve a better performance.

5 Conclusion

In this paper, we modified the traditional BABS and ACG algorithms. With the employment of demand function and overload inhibition process, the suboptimal MA algorithm was proposed to enhance the overall performance of the multiuser OFDMA systems.

The study in this paper shows that our MA algorithm could effectively reduce the total transmitting power in various cases with the constraint of BER satisfied. Via the simulation using actual services, we have proved that the performance of our MA algorithm, which is very close to the optimal MA algorithm, is 5-6 dB higher than that of the static allocation algorithms. Besides, the computing complexity of this MA algorithm is $O(K * N)$, which is much lower than that of the Wong algorithm $O(N^4)$.

However, the issues, when the CSI is imperfect, are not considered in this paper. In addition, the performance of these algorithms for OFDMA systems with heterogeneous services is still unknown. Hence, further study should be taken before this algorithm can be applied in real systems.

Acknowledgments. This work is partly supported by China's International Science and Technology Cooperation Program under Grant No.2008DFA11630, and partly by the China National Science Foundation under Grant No.60496315.

References

1. Liu, Q., Zhou, S., Giannakis, G.B.: Queuing with Adaptive Modulation and Coding over Wireless Links: Cross-Layer Analysis and Design. *IEEE Transactions on Wireless Communications* 4, 1142–1153 (2005)
2. Lee, C., Jeon, G.J.: An Efficient Adaptive Modulation Scheme for Wireless OFDM Systems. *ETRI Journal* 29, 445–451 (2007)
3. Wong, C.Y., Cheng, R.S., Lataief, K.B., Murch, R.D.: Multiuser OFDM with Adaptive Subcarrier, Bit, and Power Allocation. *IEEE Journal on Selected Areas in Communications* 17(10), 1747–1758 (1999)
4. Wong, C.Y., Tsui, C.Y., Cheng, R.S., Letaief, K.B.: A Real-time Subcarrier Allocation Scheme for Multiple Access Downlink OFDM Transmission. In: *Vehicular Technology Conference, 1999 (VTC 1999-Fall)*, IEEE VTS 50th, vol. 2, pp. 1124–1128. IEEE Press, New York (1999)
5. Yu, D.D., Cioffi, J.M.: SPC 10-2: Iterative Water-filling for Optimal Resource Allocation in OFDM Multiple-Access and Broadcast Channels. In: *Global Telecommunications Conference, 2006 (GLOBECOM 2006)*, IEEE, pp. 1–5. IEEE Press, New York (2006)
6. Zhang, Y.J., Letaief, K.B.: Multiuser Adaptive Subcarrier-and-Bit Allocation with Adaptive Cell Selection for OFDM Systems. *IEEE Transactions on Wireless Communications* 3, 1566–1575 (2004)
7. Jang, J., Lee, K.B.: Transmit Power Adaptation for Multiuser OFDM Systems. *IEEE Journal on Selected Areas in Communications* 21, 171–178 (2003)

8. Liu, G., Zhang, J., Zhou, B., Wang, W.: Scheduling Performance of Real Time Service in Multiuser OFDM System. In: International Conference on Networking and Mobile Computing in Wireless Communications, 2007 (WiCom 2007), pp. 504–507. IEEE Press, New York (2007)
9. Didem, K., Li, G., Liu, H.: Computationally Efficient Bandwidth Allocation and Power Control for OFDMA. *IEEE Transactions on Wireless Communications* 2, 1150–1158 (2003)
10. Rohling, H., Grunheid, R.: Performance Comparison of Different Multiple Access Schemes for the Downlink of an OFDM Communication System. In: IEEE Vehicular Technology Conf. in Proc (VTC 1997), Phoenix, AZ, pp. 1365–1369. IEEE Press, New York (2007)

Detecting Community Structure in Networks by Propagating Labels of Nodes

Chuanjun Pang, Fengjing Shao, Rencheng Sun, and Shujing Li

Information Engineering College of Qingdao University
sfj@qdu.edu.cn

Abstract. Community structure is a common property of many networks. Automatically finding communities in networks can provide invaluable help in understanding the structure and the functionality of networks. Many algorithms to find communities have been developed in recent years. Here we devise a new algorithm to detect communities in networks—propagation algorithm. By propagating the labels of nodes in networks, detecting communities are transformed into analyzing the labels of nodes in networks in this algorithm. Real-world and computer-generated networks are used to verify this algorithm. The results of our experiments indicate that it is sensitive to community structure and effective at discovering communities in networks.

Keywords: Complex networks, Community structure, Propagation algorithm.

1 Introduction

Complex networks, as a powerful approach to understand the complex systems, have attracted enormous amount of attentions from the scientific community. Researchers have discovered several statistical properties that many networks seem to have, such as small world property [1], scale free distribution [2,3] and so on. A property that seems to be common to many networks is community structure, the division of network nodes into groups within which the network connections are dense, but between which the networks connection are sparser, as Fig. 1 illustrates[4].

Community structure, as a common property of many networks, often hinders information and functioning of networks. For example, communities in biological may hinder information on the functioning of system. In fact, community in cellular and genetic networks may represent functional module [5], communities in World-Wide-Web may represent pages on related topics [6,7]. Therefore, detecting communities automatically in networks has practical significance to understand and analyze the functioning of systems that networks represent.

At present, researchers have proposed a lot of algorithms to discover communities in networks. Spectral bisection methods which are based on the eigenvectors of graph Laplacian are proposed in [8-10]. The hierarchical clustering algorithm [11] is presented by socialist. [4,8] describe GN algorithms which are based on edge removal. Other algorithms that are based on Modularity are proposed in [12].

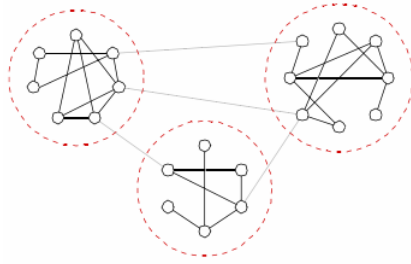


Fig. 1. A small network with community structure. In this case there are three communities, denoted by dashed circles, which have dense internal links but between which there is only a lower density of external links [4].

In this paper, we propose a new method—*propagation algorithm*, to discover community structure in networks. Experimental results show that this new algorithm is effective at detecting community structure in networks. We assume that the network is the simplest, with a single type of undirected, unweighted edge connecting unweighted vertices of a single type. The remaining of the paper is organized as follows: in section 2 concepts and terminologies referred to the algorithm are defined. In Section 3 we describe the new algorithm—*propagation algorithm*. In section 4, the effectiveness of propagation algorithm is verified by detecting the community structure in a real world network and a computer-generated network. In section 5 we summarize our study and point out the improvement of this algorithm in future.

2 Definitions

2.1 Propagation Matrix

The main idea of propagation algorithm, which comes from graph-based semi-supervised learning [15], is to propagate labels of nodes through edges by an iterative procedure. The number of intra-community edges is larger than that of inter-community edges. Hence labels travel between nodes in the same community is easier than between those in the different communities. Here we first define propagation matrix which plays an important role in this algorithm and we will propose the new algorithm and introduce the iterative procedure in next section.

Propagation matrix is a $n \times n$ matrix. \hat{T}_{ij} is the i^{th} row and j^{th} column element of matrix. The value of \hat{T}_{ij} is the fraction that the label of node i comes from node j when propagating labels and is defined as follows:

$$\hat{T}_{ij} = P(j \rightarrow i) = \frac{w_{ij}}{\sum_{k=1}^n w_{ki}}. \quad (1)$$

where n is the number of nodes in network and w_{ij} is the elements of the adjacency matrix W of the network. The matrix form of expression (1) is:

$$\hat{T} = D^{-1}W . \tag{2}$$

where D is the diagonal degree matrix and W is the adjacency matrix of networks.

In fact, propagation matrix describes how nodes affect each other when propagating labels of them.

2.2 Modularity

Modularity, which is proposed by Newman and Girvan, is a measure to value how good a particular division is [4]. For a division with g groups, define a $g \times g$ matrix e whose component e_{ij} is the fraction of edges in the original network that connect vertices in groups i to those in group j [4]. Then the modularity is defined as follows:

$$Q = \sum_i e_{ii} - \sum_{ijk} e_{ij}e_{ki} = \text{Tr}e - \|e\|^2 . \tag{3}$$

For more information about modularity, please see Ref [4, 5].

3 Propagation Algorithm

In this section, the propagation algorithm and the iterative procedure mentioned above are proposed. First of all, we find a node with maximum degree and assign its label a positive integer (in this paper, positive integers are used to represent labels of node). This node, which is the source node, propagates its label to all other nodes within the networks by an iterative procedure. Note that, within every iterative procedure the label of the source node is constant while the labels of other nodes affect each other. After the t^{th} iterative procedure, the labels of all nodes are denoted as $V^t = (v_1^t, \dots, v_i^t, \dots, v_n^t)$, n is the number of nodes and v_i^t is the label of the node i after the t^{th} iterative procedure. As mentioned above, considering the fraction that the label of node i comes from node j is \hat{T}_{ij} and the label of the node i comes from all its adjacent nodes, after the $(t+1)^{th}$ iterative procedure the label that the node i receives from node j is $\hat{T}_{ij} \times v_j^t$ and the label of node i is $v_i^{t+1} = \sum_{j=1}^n \hat{T}_{ij} \times v_j^t$. We represent the labels of all nodes after the $(t+1)^{th}$ iterative as $V^{t+1} = \hat{T} V^t$. \hat{T}_{ij} is the propagation matrix that we have defined above.

Algorithm is presented as follows:

Input: adjacent matrix of the network W .

Initialization: assume that the initial labels of nodes is $V^0 = (0, 0, \dots, k, \dots, 0)$, $v_m^0 = k$ and k is a positive integer. m is the source node (in this paper, node which has maximum degree is used as source node).

Step1: Compute the diagonal matrix D of the network.

Step2: Compute $\hat{T} = D^{-1}W$

Step3: Iterative procedure

$$V^{t+1} = \hat{T}V^t,$$

$$v_m^{t+1} = k \text{ (Set label of the source node constant)}$$

Until convergence to V^∞

Step4: determine the optimized border which divides the nodes according to the labels of nodes and the modularity.

The convergence of the iterative procedure depends on the property of the network [14]. Next section we will test propagation algorithm using a real world networks and a computer-generated networks.

4 Experiments and Results

In this section, propagation algorithm is used to discover community structure within a real world network—Karate club network. A computer-generated network is also used to test our algorithm. It is a problem when stopping the iterative procedure. The convergence of the iterative procedure has been verified in [14], so we assume that the iterative procedure stops when the average change of labels of all nodes is lower than a small positive that we set.

4.1 Karate Club Network

Karate club network is taken from one of the studies in social network analysis. Over the course of two years in the early 1970s, Wayne Zachary observed social interactions between the members of a karate club at an American university [16]. He constructed networks of ties between members of the club based on their social interactions both within the club and outside it. By chance, a dispute arose during the course of his study between the club's administer and its principal karate teacher over whether to raise club fees, and as a result the club eventually split in two smaller clubs, centered around the administer and the teacher[4]. Fig. 3(a) shows a consensus network structure extracted from Zachary's observations.

Propagation algorithm is used to detect communities in karate club networks. First of all, we find the node with maximum degree and assign its label as 1(as mentioned above, positive integers are used to represent labels of nodes, in this case we use 1 to represent label of source node), and other nodes' labels as 0. The label travel occurs in the whole network until the average change of all nodes' labels is lower to a small positive that we set. The left of Fig. 2 shows the scatter of final labels of all nodes and the right shows the correspond value of modularity when dividing the nodes at different position according to their labels. We can see that when dividing the nodes at $v=0.99927$ (dashed line) the value of modularity Q is maximum 0.3714661. At this position we divide the nodes into two parts (dashed line in the left of Fig. 2) and get the result as Fig. 3(b) shows. We can see that the algorithm classified every node into communities correctly.

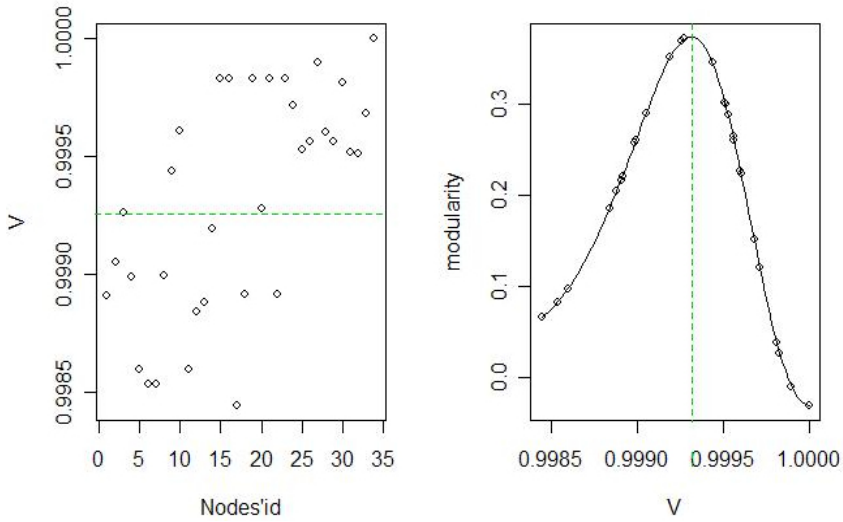


Fig. 2. Result. Left: scatter of labels of nodes in network. The horizontal axis represents the ids of nodes and the vertical axis represents the nodes' labels. Right: Correspond value of modularity when dividing the nodes at different position according to their labels.

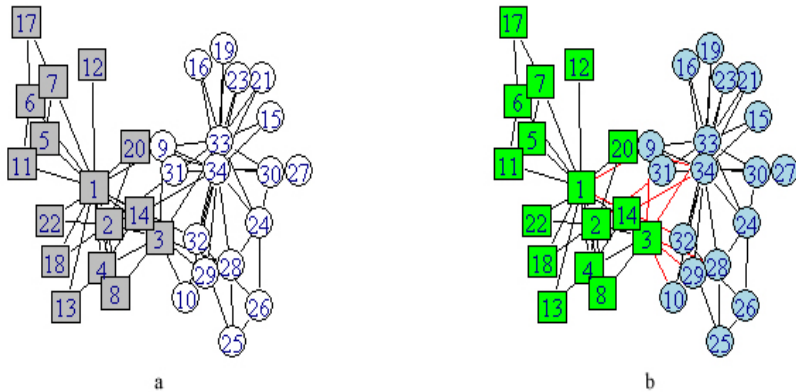


Fig. 3. Karate club network [4,16]. (a) The administrator and the teacher are represented by node 1 and 33 respectively. Shaded squares represent small club centered around the administrator and the open circles represent small club centered around the teacher. (b) The communities that detected using propagation algorithm. Every node is classified into communities correctly.

4.2 Computer-Generated Network

A network which has 128 nodes is constructed to test propagation algorithm. This network has two known communities (each community has 64 nodes) and the average degree of each node is 16. Edges are added between node pairs randomly. The

average number of inner-community edges per node is C_{out} . Fig. 4 shows the networks when $C_{out} = 2$ and $C_{out} = 8$. Obviously, as C_{out} increases, the number of inter-community edges becomes larger and the community structures within networks become ambiguous. Propagation algorithm is used to detect communities within network and we record the fraction of nodes classified correctly when we set different value of C_{out} .

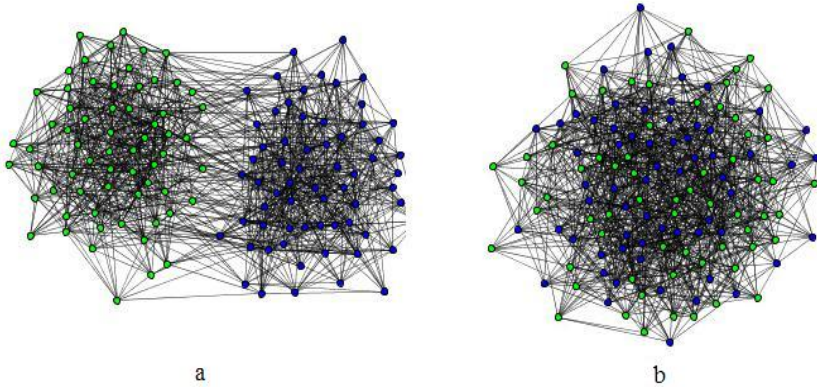


Fig. 4. Computer-generated networks. (a) $C_{out} = 2$. (b) $C_{out} = 8$.

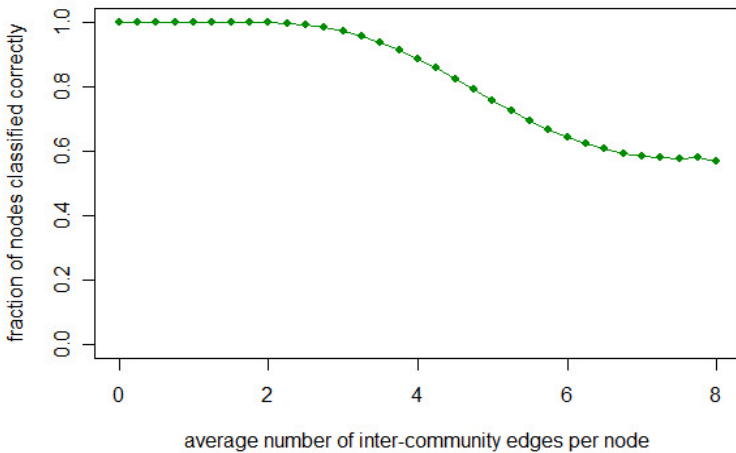


Fig. 5. Fraction of nodes classified correctly by varying C_{out} . The vertical axis represents the fraction of nodes classified correctly and the horizontal axis represents C_{out} .

In Fig. 5, we show the fraction of nodes classified in to communities correctly by varying C_{out} . We can see that, using propagation algorithm, more than 90% of all nodes can be classified correctly when $C_{out} < 4$. Although this new algorithm begins to deteriorate when $C_{out} > 4$, it can also classify about 60% of nodes correctly when $C_{out} = 8$ (At this point the average number of intra-community edges and inter-community edges is the same). In other words, our algorithm is sensitive to community structures within networks and can detect communities effectively if there are community structures within networks.

5 Conclusions

In this paper we present a new algorithm to detect communities in networks- propagation algorithm. This algorithm converts the problem of detecting communities into analyzing the labels of nodes in networks by propagating the label of source node in networks. Real-world and computer-generated network are used to test this algorithm and the results show that it is effective at discovering communities in networks.

An interesting further feature of propagation algorithm is that it can also be used to find the particular community to which a specified node belongs, without first finding all communities in the network. In some circumstance, this feature is useful. For example, in the field of web searching, one can set the node of interest as the source node and then propagate its label by propagation algorithm. One can look for the set of nodes with labels close in some sense to that of node of interest, and regard those as its community.

But one disadvantage of this new algorithm is that it only detects two communities at one time. In future, we will modify this method to make it discover more than two communities in networks. In addition, using this algorithm to analysis more complex networks, such as directed or weighted networks, is our main research direction in future work.

References

1. Milgram, S.: The Small World Problem. *Psychology Today* 2, 60–67 (1967)
2. Albert, R., Jeong, H., Barabasi, A.L.: Diameter of the World-wide Web. *Nature* 401, 130–131 (1999)
3. Barabasi, A.L., Albert, R.: Emergence of Scaling in Random Networks. *Science* 286, 509–512 (1999)
4. Newman, M.E.J., Girvan, M.: Finding and Evaluating Community Structure in Networks. *Phys. Rev. E* 69, 026113 (2004)
5. Fortunato, S., Latora, V., Marchiori, M.: Method to Find Community Structures Based on Information Centrality. *Phys. Rev. E* 70, 056104 (2004)
6. Gibson, D., Kleinberg, J., Raghavan, P.: Inferring Web Communities from Link Topology. In: *Proceedings of the 9th ACM Conference on Hypertext and Hypermedia*, Association of Computing Machinery, New York (1998)

7. Flake, G.W., Lawrence, S.R., Giles, C.L., Coetzee, F.M.: Self-organization and Identification of Web communities. *IEEE Computer* 35, 66–71 (2002)
8. Newman, M.E.J.: Detecting Community Structure in Networks. *Eur. Phys. J. B* 38, 321–330 (2004)
9. Fiedler, M.: Algebraic Connectivity of Graphs. *Czech. Math. J.* 23, 298–305 (1973)
10. Pothén, A., Simon, H., Liou, K.P.: Partitioning Sparse Matrices with Eigenvectors of Graphs. *SIAM J. Matrix Anal. Appl.* 11, 430–452 (1990)
11. Scott, J.: *Social Network Analysis: A Handbook*. Sage, London (2000)
12. Newman, M.E.J.: Fast Algorithm for Detecting Community Structure in Networks. *Phys. Rev. E* 69, 066133 (2004)
13. Wu, F., Huberman, B.A.: Finding Communities in Linear time: A Physics Approach. *Eur. Phys. J. B* 38, 331–338 (2004)
14. Olivier, C., Bernhard, S., Alexander, Z.: *Semi-Supervised Learning*, pp. 195–196 (2006)
15. Zhu, X., Ghahramani, Z.: *Learning from Labeled and Unlabeled Data with Label Propagation*. Technical Report CMU-CALD-02-107, Carnegie Mellon University, Pittsburgh (2002)
16. Zachary, W.: An Information Flow Model for Conflict and Fission in Small Groups. *Journal of Anthropological Research* 33, 452–473 (1977)
17. Girvan, M., Newman, M.: Community Structure in Social and Biological Networks. *Proc. Natl. Acad. Sci. USA* 99, 7821–7826 (2002)

Algorithm for Multi-sensor Asynchronous Track-to-Track Fusion

Cheng Cheng and Jinfeng Wang

Zhengzhou Institute of Aeronautical Industry Management
No.2, Daxuezhong Road, Zhengzhou 450015, China

Abstract. This paper derives an algorithm for multi-sensor asynchronous track-to-track fusion that combines tracks provided by different sensors, which have each communication delay. In this algorithm, an adaptive approach for fusion in a multi-sensors environment is used. The measurements of two sensors tracking the same target are processed by linear Kalman Filters, and the outputs of the local trackers are sent to the central node. In this node, a decision logic, which is based on the comparison between distance metrics and thresholds, selects the method to obtain the global estimate. The simulation result illustrates that this algorithm approaches the Weighted Covariance Fusion (WCF) algorithm in the fusion precision, and the computational burden reduces one about the half.

Keywords: Track-to-track fusion, Multi-sensors, Fusion simulation.

1 Introduction

Centralized processing architecture and distributed processing architectures are used frequently in multi-sensors data fusion system. A centralized processing architecture is often assumed in mathematically developing tracking algorithms, and it has been shown that tracking performance of centralized fusion architectures improve significantly when multiple sensors are used. However, increasing the number of sensors incurs a larger computational burden on the central processor as well as greater communication bandwidth requirements. In practice, distributed processing architectures are used due to their lower computational demands, lower communication bandwidth requirements, and greater reliability and survivability [1].

An algorithm that incorporates feedback in distributed fusion architectures for maintaining target tracks in cluttered environments is proposed. The feedback sequences are constructed by compensating global updated estimates with track information due to global predicted estimates and local track estimates. Because of its orthogonal properties, these feedback sequences are then used in the filtering process to update local estimates before local processors acquire new sets of measurements. The process of constructing these feedback sequences is presented and implemented on a proposed distributed fusion system where each local processor receives measurements from multi-sensors [2, 3]. An adaptive track fusion algorithm was developed by Beugnon [4], in which method for fusion of local track estimate is selected according to a logic-based decision tree. This article synthesized the merit of two methods

above, then derives the adaptive algorithm for multi-sensor track fusion with feedback information. The concrete computation process of the track fusion algorithm is introduced in the situation of asynchronous in detail. The simulation result illustrates that this algorithms approaches the Weighted Covariance Fusion (WCF) method in the fusion precision, and the computational burden reduces one about the half.

2 Track-to-Track Fusion Model

Consider N targets. that have been tracked by M sensors in multi-sensor data fusion system, and the state equation of target

$$X^i(k+1) = \phi_i(k+1, k)X^i(k) + W_i(k) \quad i=1, 2 \dots\dots N \quad (1)$$

where $X^i(k)$ is the state of target i, and $W_i(k)$ is a zero mean Gaussian process with covariance

$$E[W_i(k)W_i^T(j)] = Q_i \sigma_{kj} \quad (2)$$

The measurement model for target i is given by

$$Z^i(k) = H_i X^i(k) + V_i(k) \quad (3)$$

where H_i is measurement matrix, $V_i(k)$ is a zero mean Gaussian process with covariance

$$E[V_i(k)V_i^T(j)] = R_i \sigma_{kj} \quad (4)$$

The structure of the feedback fusion system is given in Fig. 1. The two sensors are assumed to have different sampling rates and different communication delays. Assume that each sensor processes its observations locally and communicates its track to a fusion center. The central processor fuses the track of both sensors and communicates the fused track back to each sensor. Each sensor implements a Kalman filter to track the target. Let $X(k, k)$ and $P(k, k)$ be the track produced by the fusion center at time $t = kT$, where T is the update period of the fused track. It is worth noting here that T can be variable to accommodate communication delays and or to allow the central computer to finish its commitment to a prior task before it can update the fused track. Let $\hat{X}_j^i(k+1, k+1)$ and $\hat{P}_j^i(k+1, k+1)$ be the track provided by sensor j , $j = 1, 2$.

At every local update period, each sensor uses the previous output of the fusion center as its initial track. In this case, at time $k+1$, the updated track of sensor j is given by

$$\begin{aligned} \hat{X}_j^i(k+1, k+1) &= \phi_i(k+1, k)X_f^i(k, k) + K_j^i[Z_j^i(k+1) \\ &\quad - H_j^i \phi_i(k+1, k)X_f^i(k, k)] \end{aligned} \quad (5)$$

$$\hat{P}_j^i(k+1, k) = \phi_i(k+1, k)P_f^i(k, k)\phi_i(k+1, k)^T + Q_j(k). \quad (6)$$

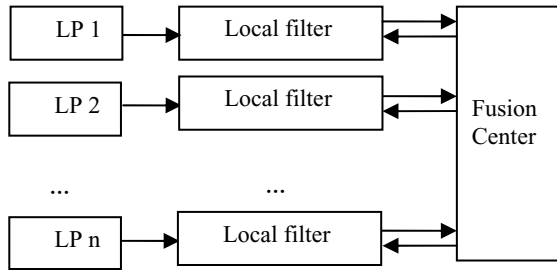


Fig. 1. Track fusion with feedback

3 Adaptive Track-to-Track Fusion Algorithm

The motivation behind adaptive track fusion comes from two observations. First, doing fusion is not always worth the required computation and bandwidth capacity. Second, the inclusion of the Cross Covariance matrix is sometimes useless and Simple Fusion may perform as well as the complete Weighted Covariance Fusion. To overcome the changes in the system characteristics and in the requirements, an adaptive approach to fusion is developed here. The sensors and the local trackers form the local nodes. The central node is divided into the decision logic, which chooses the method for evaluating the global estimates, and the fusion node, which evaluates the global estimates. The decision logic is based on the computation of two distance metrics and a simple decision tree. A first metric allows the system to choose between using the local track as the global track and performing track fusion. In the latter case, another distance is computed to decide whether to use Simple Fusion or Weighted Covariance Fusion. The architecture of the decision tree is presented in Fig (2). The local and external tracker outputs are used to compute the first distance (D1). If this distance is smaller than a given threshold (T1), the global estimate is equal to the Local Track (LT). Otherwise, a second distance (D2) is computed and compared to the second threshold (T2). Depending on this last criterion, either Simple Fusion (SF) [5] or Weighted Covariance Fusion (WCF) [6] is chosen to calculate the global estimate. The distances are functions of the tracking system and target characteristics whereas the thresholds are human inputs reflecting the trade-off between desired accuracy and computational load.

The first distance, between the local track and the fused estimate, is defined by

$$D_1 = (\hat{x}_1 - \hat{x}_{SF})^T (P_1 + P_{SF})^{-1} (\hat{x}_1 - \hat{x}_{SF}) \tag{7}$$

$$P_1 + P_{SF} = P_1(P_1 + P_2)^{-1}(P_1 + 2P_2) \tag{8}$$

$$\hat{x}_1 - \hat{x}_{SF} = P_1(P_1 + P_2)^{-1}(\hat{x}_1 - \hat{x}_2) . \tag{9}$$

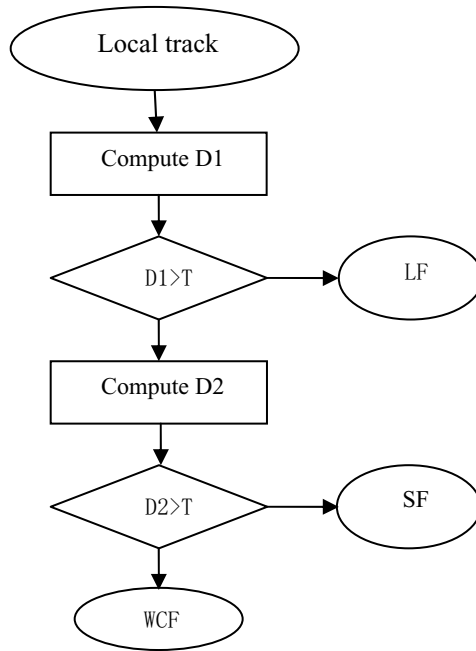


Fig. 2. Decision tree

The first distance is finally expressed by

$$D_1 = (\hat{x}_1 - \hat{x}_2)^T (P_1 + P_2)^{-1} P_1 (P_1 + 2P_2)^{-1} (\hat{x}_1 - \hat{x}_2) . \tag{10}$$

Thus, the statistics is related only to their covariance and their state estimates. Define the second distance as

$$D_2 = (\hat{x}_1 - \hat{x}_{WCF})^T (P_1 + P_{WCF})^{-1} (\hat{x}_1 - \hat{x}_{WCF}) . \tag{11}$$

By using

$$\hat{x}_1 - \hat{x}_{WCF} = (P_1 - P_{12})^{-1} P_E^{-1} (\hat{x}_1 - \hat{x}_2) \tag{12}$$

$$P_1 + P_{WCF} = (P_1 - P_{12}) P_E^{-1} P_A , \tag{13}$$

where the matrix P_A and P_E is defined by

$$P_A = P_1 + P_{21} + 2(P_2 + P_{21})(P_1 - P_{12})^{-1} P_1 \tag{14}$$

$$P_E = P_1 + P_2 - P_{12} - P_{21} , \tag{15}$$

the second distance is expressed as a function of the local agent outputs by

$$D_2 = (\hat{x}_1 - \hat{x}_2)^T P_E^{-1} (P_1 - P_{12})^{-T} P_A^{-1} (\hat{x}_1 - \hat{x}_2). \tag{16}$$

4 Apply the Algorithm to the Asynchronous Sensor System

After we have obtained the adapted fusion algorithm, we were allowed to use it to obtain the fused track. When using equation (5) and (6) to compute, we have supposed each local sensor is synchronization, and the fusion center also is synchronization. But in fact, the sensor is the asynchronous, the sampling speed and the correspondence delay is asynchronous also [7]. Therefore, when carrying on the information fusion with feedback structure in the engineering application, we must consider the question of sensor asynchronous. In order to analyze conveniently, the system is supposed has two asynchronous sensors (LP1 and LP2). Fig. 3 illustrated the succession relations of partial sensor and the fusion center fusion.

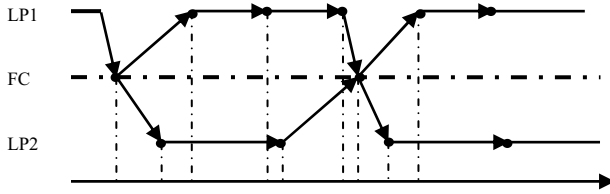


Fig. 3. Asynchronous track fusion succession

Using the algorithm introduced above, it is very easy to obtain the state estimation of sensor LP1

$$\hat{X}_1(t_{m+1}^1, t_{m+1}^1) = \phi_1(t_{m+1}^1, T_k) X_f(T_k, T_k) + K_1 [Z_1(t_{m+1}^1) - H_1 \phi_1(t_{m+1}^1, T_k) X_f(T_k, T_k)] \tag{17}$$

where the matrix $\phi_1(t_{m+1}^1, T_k)$ is defined by

$$\phi_1(t_{m+1}^1, T_k) = e^{A(t_{m+1}^1 - T_k)}. \tag{18}$$

Using equation (6), the covariance matrix of LP1 may be obtained

$$\hat{P}_1(t_{m+1}^1, T_k) = \phi_1(t_{m+1}^1, T_k) P_f(T_k, T_k) \phi_1(t_{m+1}^1, T_k)^T + Q_1(T_k). \tag{19}$$

In turn, we may obtain the state estimations and the covariance matrix of sensor 2.

5 Performance Evaluation

In the simulations, the distributed system which has two sensors with feedback architecture tracks two independent targets moving. The model used to track the target is a CA model for all filters. The measurement errors in position and in angle of local sensor 1 are $\sigma_{r1}=400\text{m}$, $\sigma_{\theta1}=0.02\text{rad}$, sampling interval is $T_{m1}=1.5\text{s}$; The measurement errors in position and in angle of local sensor 2 are $\sigma_{r2}=200\text{m}$, $\sigma_{\theta2}=0.015\text{rad}$, sampling interval is $T_{m2}=0.8\text{s}$; The goal is tracked holds 60 seconds, and this simulation consists of 100 Monte Carlo runs.

Fig. 4 gives the position RMS error and compares the tracking performance with and without fusion system. The fused precision is good to any local sensor, and this has been conforming to the aspect of multi-sensors data fusion in the target tracking. The speed RMS error and performance comparison given by Fig. 5 also have the same conclusion.

Fig. 6 compares the fusion performance using SF algorithm, WCF algorithm and the algorithm N derived in this paper. Because WCF algorithm has considered the covariance between local track, the fusion performance is excelled the SF algorithm.

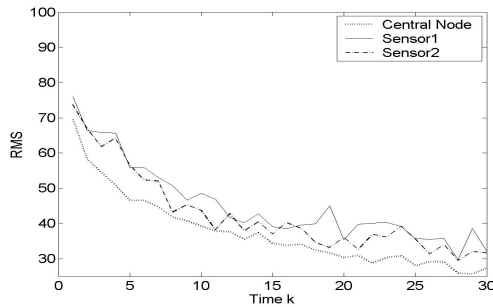


Fig. 4. Position RMS error of tracking target 1

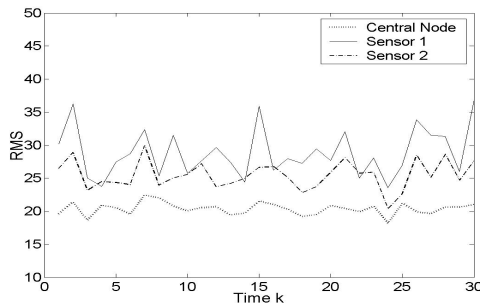


Fig. 5. Speed RMS error of tracking target 1

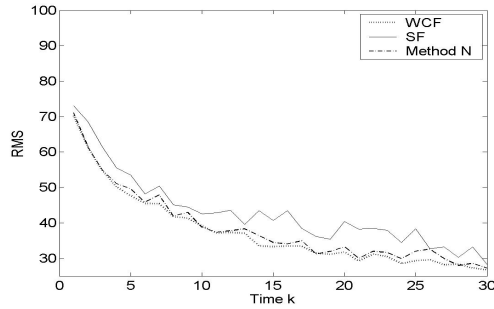


Fig. 6. Position RMS error of tracking target 2

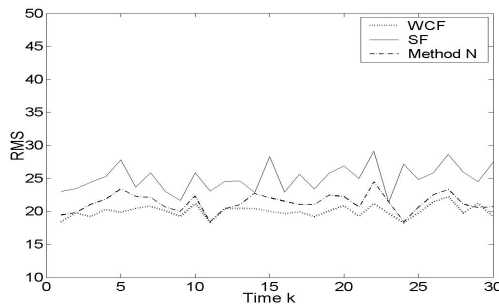


Fig. 7. Speed RMS error of tracking target 2

Table 1. Fusion time

Fusion Method	Times(s)
SF	48
WCF	125
Adaptive	60
Method of this paper	66

The speed RMS error and performance comparison given by Fig. 7 also have the same conclusion. But, the computational burden increases very much. Algorithm N has used the auto-adapted technology, therefore this algorithm approaches the WCF method in the fusion precision, and the computational burden reduces much. This is illustrated by Table 1.

6 Conclusion

In this paper, the multi-sensors data fusion system with distributional structure has been introduced. The feedback architecture may enhance the tracking performance of local sensors, therefore the precision of center fusion is increased. The adaptive fusion algorithm may satisfy the characteristic that the performance of fusion system may be

changed unceasingly, and balances the contradiction between the fusion precision and computational burden reduces. This article synthesized the merit of two methods above, and derives the adaptive algorithm for multi-sensors track fusion with feedback information. The concrete computation process of the track fusion algorithm is introduced in the situation of asynchronous in detail. The simulation result illustrates that this algorithms approaches the WCF method in the fusion precision, and the computational burden reduces one about the half.

References

1. He, Y., Xiu, J.: Radar Data Processing with Applications. Publishing House of Electronics Industry, Peking (2006)
2. Drummond, O.E.: Feedback in Track Fusion without Process Noise. In: Proc. SPIE Conf. Signal Data Process, Small Targets, vol. 2561, pp. 147–152 (1995)
3. Alouani, A.T., Rice, T.R.: Asynchronous Track Fusion Revisited. IEEE Transactions on Aerospace and Electronic Systems 1021, 118–122 (2005)
4. Malmberg, A., Karlsson, M.: Track-to-Track Association in Decentralized Tracking System with Feedback. In: Proc. SPIE Conf. Signal Data Process, Small Targets, vol. 4048, pp. 461–472 (2000)
5. Singer, R.A.: Estimating Optimal Tracking Filter Performance for Manned Maneuvering Targets. IEEE Transactions on Aerospace and Electronic Systems 72, 473–483 (1970)
6. Bar-Shalom, Y.: On the Track-to-Track Correlation Problem. IEEE Transactions on Automatic Control 26, 571–572 (1981)
7. Yang, W.: Multisensor Data Fusion with Application. Xidian University Press, Xi'an (2004)

Remote Sensing Based on Neural Networks Model for Hydrocarbon Potentials Evaluation in Northeast China

Shengbo Chen

Geoexploration Science and Technology, Jilin University, 130026, Changchun, China
chensb@jlu.edu.cn

Abstract. Hydrocarbon resources shortage is a wide-world issue, which causes great efforts being made for hydrocarbon resources exploration all over the world. Songliao Basin is one of the most important potential regions for hydrocarbon resources in Northeast China. Owing to cost saving of hydrocarbon exploration, it is necessary to evaluate the regional potential of hydrocarbon by remote sensing before costly hydrocarbon exploration and drilling. In the study, Landsat TM data at the western slope of Songliao Basin are processed to improve hydrocarbon-related linear-circular structures and micro-seepages information. A self-organizing neural network is built for the evaluation to hydrocarbon potentials of unknown areas. Twelve features are integrated into the model from remote sensing, geophysical anomaly, and geological setting around the western slope of Songliao Basin. The model is trained by a competitive learning of twelve features of four hydrocarbon-known boreholes. The hydrocarbon potentials in three unknown circular clusters are evaluated successfully by the trained self-organizing neural networks model.

Keywords: Songliao basin, Landsat TM, Self-organizing neural network, Hydrocarbon potentials.

1 Introduction

Hydrocarbon resources shortage is a wide-world issue, which causes great efforts being made for hydrocarbon resources exploration all over the world. Songliao Basin is one of the most important potential regions for hydrocarbon resources in Northeast China. Owing to cost saving of hydrocarbon exploration, it is necessary to evaluate the regional potential of hydrocarbon by remote sensing before costly hydrocarbon exploration and drilling.

The application of remote sensing, especially the mapping of a regional structure, has a very long tradition worldwide[1]. Geological structures on remotely sensed images can also be enhanced by using different directional filters, principal component analysis (PCA), and intensity-hue-saturation (HIS)[2]. These structures are important for hydrocarbon-bearing characteristics of geologic traps[3]. Landsat Thematic Mapper (TM) thermal data (band 6) have been taken as a recognition element to map distribution of hydrocarbon [4]. Reflectance in the visible and near-infrared wavelengths provides a rapid and inexpensive means for determining the mineralogy of samples and obtaining information on chemical composition. Hydrocarbon micro-seepage theory

establishes a relation between hydrocarbon reservoirs and some special surface anomalies, which mainly including surface hydrocarbon micro-seepage and related alterations. Some methods have been explored for oil-gas determination by the reflectance spectra of surface anomalies[5,6]. Therefore, the structure traps and micro-seepage from remote sensing image can provide important and indirect information before hydrocarbon exploration.

Over the last decades, the application of neural networks for the hydrocarbon evaluation has been a subject of research. According to the geological model, an artificial neural network with three-layer networks, multi-input, and single output, is built to evaluate the hydrocarbon generation condition of the Ordovician System in Ordos Basin. The result is basically the same as that of the expert system [7]. Some neural networks, including self-organizing neural network and fuzzy neural network, were used for hydrocarbon prediction. With the neural network method, geochemical exploration, geoelectric, and remote sensing were integrated to evaluate hydrocarbon-bearing characteristics of Liandaowan area, with satisfactory results obtained[8-10].

By hydrocarbon related features extraction from Landsat TM data, a self-organization neural network is explored to integrate geophysical and geological information of the study area for evaluation of hydrocarbon potential in Western Slope of Songliao Basin, Northeast China.

2 Study Area

The Songliao Basin is located in northeast China, where oil production has made up a large portion of the national supply for nearly 45 years. The boundaries of the Songliao

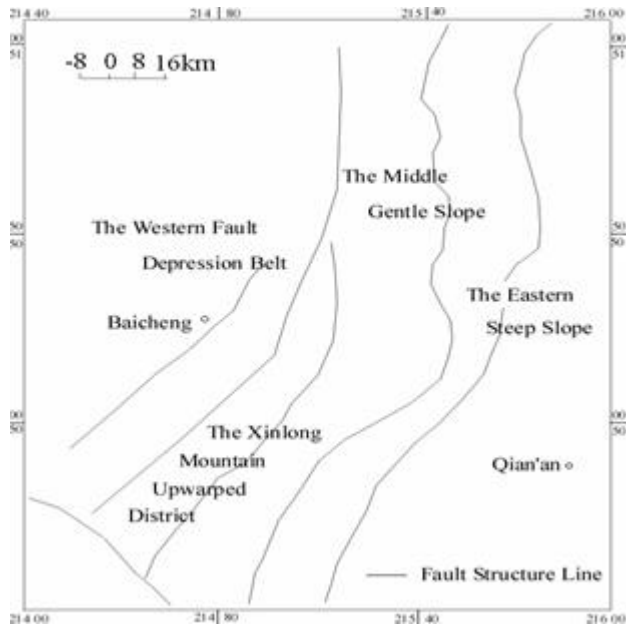


Fig. 1. Sketch Map of Geological Structural Setting in Western Slope of Songliao Basin, Northeast China

Basin are outlined in Fig.1. The basement rocks, including the carbonate rock series, gneiss, schist, and granitoid from the exposures and boreholes, are distributed under a thick sedimentary cover from the Mesozoic and Quaternary eras. The basement faults are developed into four fault systems, trending along NNE-NE, NW, EW, and NNW.

The Western slope of Songliao Basin, near to the city of Baicheng (Fig.1), slopes eastwardly with a largest depth of 2400m in the east of the slope. As a potential region for hydrocarbon resources, the study area is classified structurally into four parts: the western fault depression belt, Xinlong Mountain uplift block, the middle gentle slope and the eastern steep slope. The surface faults, stretching along NE, NNE, NW, NNW, and EW directions, are dominated and influenced by the basement faults. Thus, the faults in the study area vary in strike and depth, and play different roles in hydrocarbon displacement and conservation. It is an essential task to investigate the depths and strikes of geological structures in our oil and gas exploration mission. Meanwhile, the circular structures play sometimes an important role for the trapping of hydrocarbon.

3 Feature Extraction

Landsat TM data are processed and interpreted to produce hydrocarbon related information, including the regional distribution of geological structures, oil-gas microseepage, and thermal anomaly in the study area. All image processing was undertaken in the commerce software PCI EASI/PACE image processing software and MapGIS GIS software.

3.1 Image Pre-processing

Two scenes of Landsat TM data, path/row 120/28 and 120/29, acquired on 28 October 1989, were color matched and used to create a seamless mosaic of the entire study area. The spatial resolution and pixel spacing are both 30m. It is necessary to acquire several scenes of Landsat TM to cover the large geographical area for studying the regional structures.

Geometric correction was used to orient the satellite image to a suitable map projection, with the conventional X, Y polynomial-based transformation. Here the Universal Transverse Mercator (UTM) coordinate system with a false central meridian at 123 degree was used. Nine ground control points were defined from the topographical map at scale 1:50000. Then the subset was generated by masking the study area boundary with the Landsat TM image.

3.2 Image Enhancement

PCA undertakes a linear transformation of a set of numerical variables to create a new variable set with principal component (PC) that are uncorrelated[11,12]. A false color composite is made by the PC1 from TM5-TM7(R), TM4 (G), and PC1 from TM1-TM2-TM3 (B) (Fig.2). By comparison, this RGB image is more suitable to be interpreted for the geological structures.

In order to enhance the linear features, three different directional filtrations, including NE, NW, EW, were accomplished separately. Three filtering modules follow as

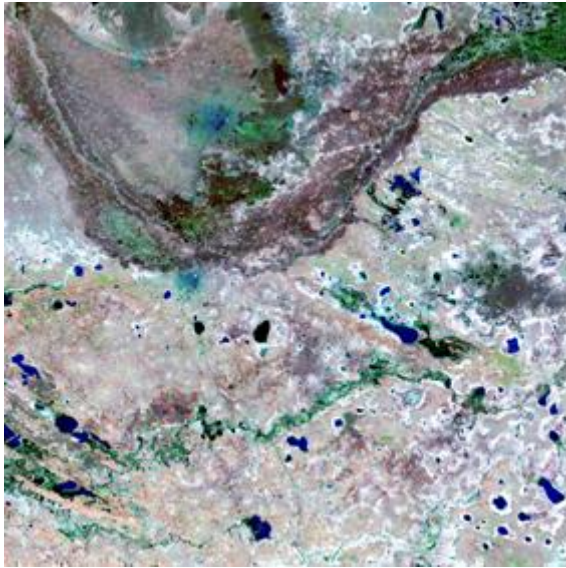


Fig. 2. False color composite image made from PCI results of Landsat TM: PC1 from TM5-TM7 (R), TM4 (G), and PC1 from TM1-TM2-TM3 (B)

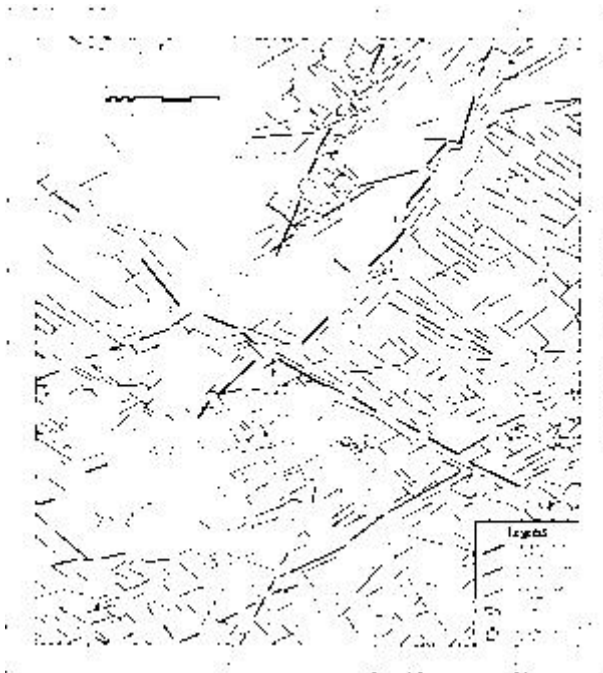


Fig. 3. Lineament map of study area from their directional filtered images. The heavy lines describe the regional sizes and clear indications.

Fig.2 respectively. The image has highlighted the major lineaments trending NW, NNW, NE, NNE, and EW. Based on Fig.2, and combining with the filtered images at three directions as well, all the lineaments, thus, are identified carefully by their linear features, including tones, and drainage system (abnormal portions of rivers, linear borders of water body, and lines of small lakes), some of which are of regional size and clear indication, shown in heavy lines (Fig.3). On Fig.3, the different size and indication of the linear features are delineated by different weights, for example, those lineaments trending NW, NE, NNE, and EW in heavy lines play significant roles in the study area.

3.3 Hydrocarbon Micro-seepage

Based on the spectral features of oil-gas polluted soil and Landsat TM band, a method was explored to improve the hydrocarbon micro-seepage information. The ratios of TM4/TM7 (R), TM4/TM5 (G), and TM4/TM3 (B), are composite into color images with the color anomalies of hydrocarbon micro-seepage.

4 Neural Networks Model

The neural network consists of artificial neurons or units, which are inter-connected with each other and arranged into three layers: an input layer, a hidden layer, and an output layer. Self-organizing in networks is one of the most fascinating topics in the neural network field. Such networks can learn to detect regularities and correlations in their input and adapt their future responses to that input accordingly. The neurons of competitive networks learn to recognize groups of similar input vectors. All the work on self-organizing neural networks was completed under the commence software Matlab 6.5.

4.1 Competitive Learning

The neurons in a competitive layer distribute themselves to recognize frequently presented input vectors. The architecture for a competitive network is shown (Fig.4). The $\| \text{dist} \|$ box in this figure accepts the input vector p and the input weight matrix $IW_{1,1}$, and produces a vector having S^1 elements. The elements are the negative of the distances between the input vector and vectors i $IW_{1,1}$ formed from the rows of the input weight matrix. Compute the net input n^1 of a competitive layer by finding the negative distance between input vector p and the weight vectors and adding the biases b . If all biases are zero, the maximum net input a neuron can have is 0. This occurs when the input vector p equals that neuron's weight vector.

The competitive transfer function accepts a net input vector for a layer and returns neuron outputs of 0 for all neurons except for the winner, the neuron associated with the most positive element of net input n^1 . The winner's output is 1. If all biases are 0, then the neuron whose weight vector is the closest to the input vector has the least negative net input and, therefore, wins the competition to output a^1 .

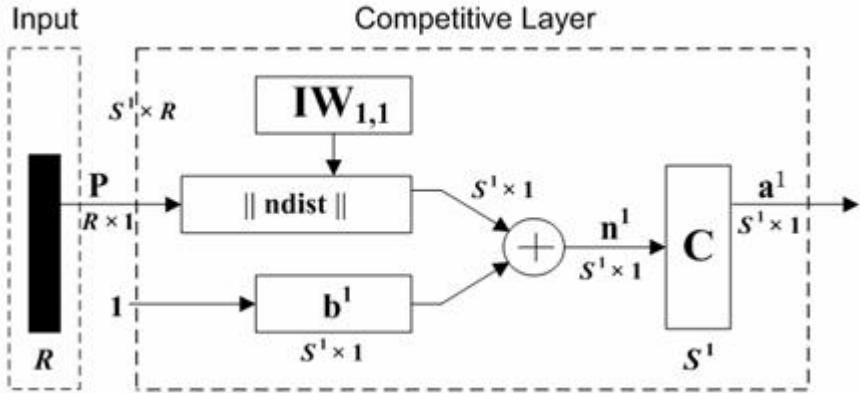


Fig. 4. The architecture for a competitive network. The relationship among the input layer, input weight, competitive layer, and output vector are generated

4.2 Training and Classification

Each neuron competes to respond to an input vector \mathbf{p} . If the biases are all 0, the neuron whose weight vector is the closest to \mathbf{p} gets the highest net input and, therefore, wins the competition and outputs 1. All other neurons output 0. The weights of the winning neuron (a row of the input weight matrix) are adjusted with the Kohonen learning rule. Supposing that the i th neuron wins, the elements of the i th row of the input weight matrix are adjusted as shown below

$${}_i\mathbf{IW}^{1,1}(q) = {}_i\mathbf{IW}^{1,1}(q-1) + \alpha(\mathbf{p}(q) - {}_i\mathbf{IW}^{1,1}(q-1)) \tag{1}$$

The Kohonen rule allows the weights of a neuron to learn an input vector, and it is useful in recognition applications. Thus, the neuron whose weight vector was the closest to the input vector is updated to be even closer. The result is that the winning neuron is more likely to win the competition, for the next time a similar vector is presented, and less likely to win when a very different input vector is presented. As more and more inputs are presented, each neuron in the layer closest to a group of input vectors soon adjusts its weight vector toward those input vectors. Eventually, if there are enough neurons, every cluster of similar input vectors will have a neuron that outputs 1 when a vector in the cluster is presented, while outputting a 0 at all other times. Thus, the competitive network learns to categorize the input vectors it sees.

5 Results

The hydrocarbon related features are derived from Landsat TM (Fig.5). These features comprise of linear structures, circular structures, and hydrocarbon microseepage in the western slope of Songliao Basin. The circular structures clusters are labeled as C1, C2, C3, C4, C5, respectively, from north to south in the study area. There are six boreholes (T4, T5, T6, T13, B52, B54), distributed in or around

the circular structures. The hydrocarbon cannot be found in the borehole T4, and the potentials in the boreholes T6, T5, and T13, are becoming better and better.

Twelve features will be selected for the boreholes (T4, T5, T6, and T13) and circular structures clusters (C3, C4, and C5) to learn and train the neural network networks. They are related to the potential of hydrocarbon in the study, such as huge and small circular structures, NE linear structures, NW linear structures, NNW linear structures, NEE linear structures, gravity anomalies, geo-magnetic anomalies, the alteration anomalies of hydrocarbon microseepage, the brightness difference of image, thermal anomalies, and heavy oil revealing on land surface.

The self-organizing neural networks are initialized by the input vector matrix of twelve features and the numbers of neuron. The training are conducted by the network simulation and learning. And the winner of neuron will produce its weights. The potential of four boreholes, T5, T6, T13, are T4, are taken as the known samples and classified as four groups according to their hydrocarbon status respectively. Then the unknown samples, including the circular structure clusters, C3, C4, and C5, three circular structure clusters in the south in Fig.5, will be evaluated by the trained networks. The results show that these circular structures and the known boreholes will be classified into three groups: C3 and T13, C4 and T6, C5 and T5.

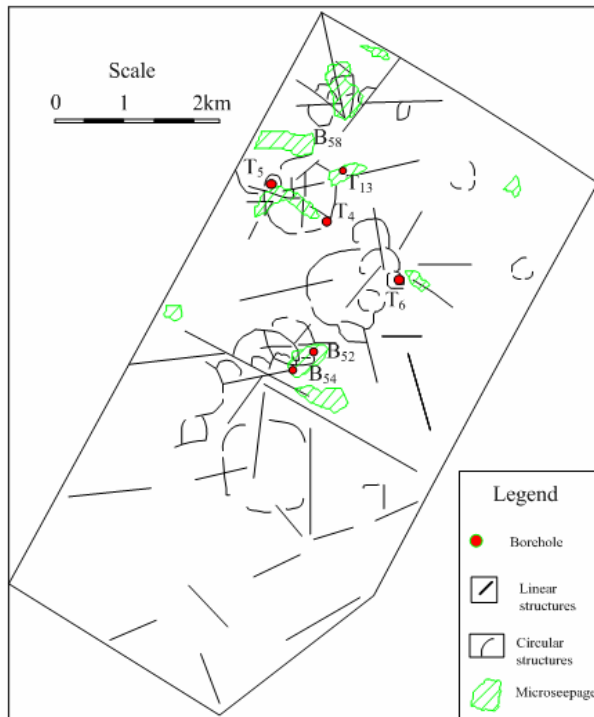


Fig. 5. Features extraction from Landsat TM. The features involving in linear structures, circular structures, and hydrocarbon microseepage.

Therefore, based on the potential status of three boreholes (T5, T6, and T13), three circular structure clusters (C3, C4, and C5), indicate good potential of hydrocarbon. Furthermore, there will be the best potential of oil-gas for the big circular structure in the southmost area in Fig.5 (C5), and the circular cluster C4 (next to C5, including boreholes B52 and B54) is better than C3 (including boreholes T6) in the potential of hydrocarbon. It is valuable to have the hydrocarbon exploration for these areas.

6 Conclusions

Remote sensing is a rapid and cost-saving technique for hydrocarbon related information extraction. In the study, based on Landsat TM data, the features, including structures and hydrocarbon micro-seepage information, are achieved in the western slope of Songliao Basin by the image processing.

With the help of self-organizing neural networks model of Matlab, twelve features are integrated for the evaluation of hydrocarbon potential in the study area. By the competitive learning and training of four known boreholes, three unknown circular structures groups are classified successfully. They indicate good potential of hydrocarbon, especially in C5. Obviously, it is valuable to have a remote sensed based neural network model for the evaluation of hydrocarbon potential before the costly hydrocarbon exploration.

References

1. Chen, S., Zhou, Y.: Classified Depth-layered Geological Structures on Landsat TM Images by Gravity Data: a Case Study of The Western Slope of Songliao Basin, Northeast China. *J. International Journal of Remote Sensing* 26, 2741–2754 (2005)
2. Ricchetti, E.: Structural Geological Study of Southern Apennine (Italy) Using Landsat 7 Imagery. *J. International Geoscience and Remote Sensing Symposium (IGARSS)* 1, 211–213 (2002)
3. Liu, T., Shi, X., Hou, L., Yuan, H.: Evaluating The Hydrocarbon-bearing Characteristics of Geologic Traps in Yakela Area in Northern Tarim Basin by Comprehensive Pattern Recognition Technique. *J. Oil Geophysical Prospecting* 32, 556–564 (1997)
4. Uddin, S., Dousari, A., Literathy P.: Evidence of Hydrocarbon Contamination from The Burgan Oil Field, Kuwait-Interpretations from Thermal Remote Sensing Data. *Journal of Environment Management* 86, 605–615 (2008)
5. Werff, H., Bakker, W., et al.: Combing Spectral Signals and Spatial Patterns Using Multiple Hough Transformations: An Application for Detection of Natural Gas Seepages. *J. Computers & Geosciences* 32, 1334–1343 (2006)
6. Xu, D., Ni, G., Jiang, L., et al.: Exploring for Natural Gas Using Reflectance Spectra of Surface Soils. *J. Advances in Space Research* 41, 1800–1817 (2007)
7. Sun, S., Wu, Z., Bei, F.: Application of an Artificial Neural Network to Evaluating Conditions of Hydrocarbon Generation. *J. Oil & Gas Geology* 17, 68–70 (1996)
8. Zhang, X., Li, Y., Liu, H.: Hydrocarbon Prediction Using Dual Neural Network. *J. SEG annual meeting* 70, 1440–1443 (2000)
9. Xiong, Y., Bao, J., Xiao, C.: Hydrocarbon Reservoir Prediction Using Fuzzy Neural Network. *J. Oil Geophysical Prospecting* 35, 222–227 (2000)

10. Doraisamy, H., Vice, D., Halleck, P.: Detection of Hydrocarbon Reservoir Boundaries Using Neural Network Analysis of Surface Geochemical Data. *J. AAPG Bulletin* 84, 1893–1904 (2000)
11. Kenea, N.H.: Digital Enhancement of Landsat Data, Spectrsal Analysis and GIS Data Integration for Geological Studies of The Derudeb Area, Southern Ren Sea Hills, NE Sudan. *Berliner Geowissenschaftliche Abhandlungen(D)* 14, 111–116 (1997)
12. Ricotta, C., Avena, G.C., Marchetti, M.: The Flaming Sandpile: Self-organized Criticality and Wildfires. *J. Ecol. Model* 119, 73–77 (1999)

A Multiple Weighting Matrices Selection Scheme Based on Orthogonal Random Beamforming for MIMO Downlink System

Li Tan, Gang Su^{*}, Guangxi Zhu, and Peng Shang

Department of Electronics & Information Engineering
Huazhong University of Science & Technology, Wuhan 430074, P.R.China
ltan@mail.hust.edu.cn, gsu@mail.hust.edu.cn,
gxzhu@mail.hust.edu.cn, sp158@smail.hust.edu.cn

Abstract. In this paper, a multiple weighting matrices selection scheme based on orthogonal random beamforming is discussed to increase the throughput of the multi-user MIMO systems. Multiple weighting matrices are performed and computed during the mini-slots. And the base station selects the best matrix based on the partial channel state information and performs the orthogonal random beamforming during the time slot. The proposed scheme over multi-user random beamforming can increase the throughput of multi-user MIMO system significantly, and the simulation results show the validity.

Keywords: MIMO system, Orthogonal random beamforming, Multiple weighting matrices selection.

1 Introduction

With the development of mobile multimedia service, the future wireless communication system calls for quick transmit speed and high quality reliability. All of them are based on the increase of the throughput and of the good spectrum efficiency. Especially, the traffic demand in the downlink is generally much higher than that of the uplink.

In multi-input multi-output (MIMO) system the downlink broadcast (BC) channel can be separated into multiple sub-channels by using space-time codes, and the base station (BS) can communicate with multiple subscribers at the same time slot, so the throughput of the system increases distinctly. This is called the multiple transmit gain for the downlink BC channel. The sum rate capacity of MIMO downlink BC channel was analyzed[1-3] and the dirty paper coding (DPC) was shown to be the optimal capability-achieving scheme[1, 4].

Although DPC is optimal in exploiting the multi-user diversity in the MIMO system, the computing complexity is unacceptable and the assumption that the BS knows all the channel state information (CSI) to every subscriber is not reasonable for the mobility of subscribers and for the limited bandwidth of the feedback channel.

^{*} Corresponding author.

Pramod Viswanath and David N. C. Tse propose a random beamforming (RBF) scheme[1] to exploit the multi-user diversity of the multi-user MIMO system. RBF scheme is simple and practical. By using a random vector with each sub-channel, RBF scheme induces large and fast channel fluctuation, even though the subscriber is in little scattering and/or slow fading environments. Then the BS finds out the perfect one to transmit, and additionally, the fairness among subscribers is concerned. RBF only needs the partial CSI to be feed back. Masoud Sharif proposed an orthogonal random beamforming (ORBF) scheme[5] to service multiple subscribers in each time slot by forming multi-beam. When the subscriber number is large, ORBF is shown to yield the optimal capacity scaling law of $M \log \log K$, which is the optimal capability of DPC[5].

By exploiting particular features of the channels, more effective methods are proposed to improve the throughput. Additional CSI of the selected subscriber group is used for optimizing the transmit power allocation[6]. Time coherence of slow fading channels is exploited by using the chosen antenna gain configuration for the scheduled user[7]. Literature[8] shows an scheme where the scheduling stage for multi-user MIMO system is aided by forming a channel vector estimate based on the combination of transmit correlation matrix and instantaneous beam Signal-to-Interference-plus-Noise-Ratios ($SINR$) feedback. A random matrix and updated feedback-aided beamforming matrices are used to increase the throughput of the ORBF system[9]. The scheme exploits the memory of the time correlation channel.

All above schemes are based on the idea of RBF and ORBF [1, 5]. Especially, only one weighting matrix is performed in each time slot. In Literature[10] Multiple random weighting vectors are performed at each mini-slot, and the BS finds the best random weighting vector and schedules the best subscriber, this scheme focuses on the case of single beam is performed only one subscriber can be serviced at each time slot. The OSDMA-S (OSDMA system with beam selection) scheme[11] uses multiple unitary matrixes in a time slot and selects the best beam group to service multiple users. The performance of OSDMA-S scheme and the optimal number of the mini-slot are analyzed and derived by the authors. But the fairness between users are not concerned. In this paper, we combine the multiple weighting matrix selection scheme (MWMS) with the proportional fairness scheme (PFS) to achieve the fairness of system. The MWMS-PFS scheme is also based on the ORBF and OSDMA-S and the fairness between users are considered. Additionally, it requires only little feedback from the subscribers (in the form of individual $SINR$ s).

This paper is organized as follows. In section 2, the system model is presented. In section 3, we review the capability of the MIMO downlink BC channel. The MWMS scheme is presented and analyzed in section 4. Simulation results are provided in section 5 to illustrate the validity of the MWMS scheme.

2 System Model

The system model is similar with [5], We consider the downlink BC channel of a multi-user MIMO system. The BS is equipped with N_t antennae, and each subscriber with N_r antennae, while K subscribers are uniformly distributed in the cell. The BS

generates the unitary beamforming matrix $\mathbf{Q}(t) \in \mathbb{C}^{N_t \times N_t}$ according to an isotropic distribution[12] and forms $B(\leq N_t)$ orthogonal random beams using the vector $\mathbf{q}_m(t)$, which is the m^{th} column of $\mathbf{Q}(t)$, and services B subscribers in one time slot. The received signal of the k^{th} subscriber at time slot t is mathematically described as

$$y_k[t] = \sum_{i=1}^B \mathbf{h}_k[t] \mathbf{q}_i[t] s_i[t] + w_k[t] \quad k = 1, 2, \dots, K, \tag{1}$$

where the $\mathbf{s}[t] \in \mathbb{C}^{N_t \times 1}$ is the pilot vector signal, $\mathbf{h}_k[t]$ is the complex channel vector of the k^{th} subscriber, $\mathbf{w}_k[t] \in \mathbb{C}^{N_t \times 1}$ is complex AWGN vector. Each element of $\mathbf{h}_k[t]$ and $\mathbf{w}_k[t]$ is independent and identically distributed (i.i.d.) with zero mean and unit variance. In this paper the time-varying Rayleigh fading channels are considered. And the subscribers know the CSI matrix \mathbf{h}_k perfectly. The total transmit power P is $\text{Tr}(E[\mathbf{s}\mathbf{s}^H])$, and averagely distributed to the transmit antennae. Here we assume $\text{Tr}(E[\mathbf{s}\mathbf{s}^H])$ is equal to N_t , thus the transmit power per antenna is 1. Without loss of generality, we consider $N_r = 1$ in the following.

so the SINR of the k^{th} subscriber over the m^{th} beam can be computed as:

$$\text{SINR}_{k,m} = \frac{|\mathbf{h}_k \mathbf{q}_m|^2}{1 + \sum_{i=1, i \neq m}^B |\mathbf{h}_k \mathbf{q}_i|^2}, \tag{2}$$

where $k = 1, 2, \dots, K$, $m = 1, 2, \dots, N_t$. Then each subscriber feeds back the maximum value $\text{SINR}_{k,m}$ and the corresponding beam index m . The BS assigns the subscriber with $k^* = \arg \max_{k=1, 2, \dots, K} (\text{SINR}_{k,m})$ for the m^{th} beam according to the partial CSI. And then BS multiplies the information symbol with the beamforming vector \mathbf{q}_m and transmits to the k^{th} subscriber.

3 MWMS Scheme

With the assume of all CSI can be feedback to the BS and no bit error occurs during the transmit, paper [1, 2] analyzed the capability of MIMO BC channel, and the sum rate capacity is

$$C_{\text{sum}} = E\left\{ \max_{P_1, \dots, P_k, \sum_k P_k = P} \log \det(\mathbf{I} + \sum_{k=1}^K \mathbf{h}_k^* \mathbf{P}_k \mathbf{h}_k) \right\}, \tag{3}$$

where $\{P_1, P_2, \dots, P_k\}$ is the optimal power allocation, and P is the total transmit power[1, 5]. DPC is the optimal scheme to get the asymptotic capability of equation (3). But the complexity of DPC makes it not realizable. ORBF scheme[5] is examined

to be the practical strategy and it can also have the same scaling laws. ORBF scheme only requires that each subscriber feeds back the maximum value $SINR_{k,m}$ and the beam index m , so the system become more practical and the uplink channel becomes more efficiency. When subscriber number is large, the sum rate capacity of system has the scaling law of $N_t \log \log K$, which is the same as DPC. But when the subscriber number is comparable with the number of transmit antennae, the gap between the ORBF and DPC is distinct. The reason is ORBF scheme generates the unitary matrix \mathbf{Q} randomly, the beamforming vector \mathbf{q}_m may not match the CSI matrix very well. And then the throughput of system is limited.

3.1 Scheme Description

We propose a multiple weighting matrices selection scheme (MWMS) to improve the performance of the conventional ORBF. The main idea of MWMS scheme is that multiple weighting unitary matrices are generated and computed during the mini-slot, the BS finds out a perfect unitary matrix to perform multi-beam during the time slot.

In MWMS scheme, the BS generates L unitary matrices $\mathcal{Q} = \{Q^1, Q^2, \dots, Q^L\}$ with $Q^l \in \mathcal{U}(N_t, N_t)$ according to an isotropic distribution at each time slot. The set \mathcal{Q} is variable between time slots. We assume that the downlink BC channel is stationary in one time slot and variable from one to another. There are L mini-slots in each time slot, the BS using the unitary matrix Q^l for the l^{th} mini-slot. Each subscriber computes the $SINR$ over each beam and feeds back the maximum value of $SINR$ and the corresponding beam index m . The BS estimates the sum rate capacity of each weighting unitary matrix according to the feedback partial CSI and schedules the subscribers over each beam. And then the BS performs the multi-beam and transmits the data information.

The operations of MWMS scheme is show as:

During the time slot t

Step 1: The BS generates $\mathcal{Q} = \{Q^1, Q^2, \dots, Q^L\}$ and uses Q^l for the l^{th} mini-slot.

Step 2: Each subscriber computes the $SINR_{k,m}^l$ and feeds back the maximum $SINR_{k,m}^l$ and the beam index m

Step 3: The BS estimates the sum rate capacity of each Q^l and finds out Q^{l^*} with the maximum sum rate capacity $C_{sum}^{l^*}$.

Step 4: The BS performs the Q^{l^*} for the time slot.

3.2 Performance Analysis

To analyze the performance of MWMS scheme, we assume the system model described in the section 2. The BS generates $\mathcal{Q} = \{Q^1, Q^2, \dots, Q^L\}$ with $Q^l \in \mathcal{U}(N_t, N_t)$ for each time slot. In the l^{th} mini-slot, the BS forms B orthogonal

random beams and services B subscribers using the unitary matrix Q^l , so the received signal of the k^{th} subscriber in the l^{th} mini-slot is

$$y_k^l = \sum_{i=1}^B \mathbf{h}_k \mathbf{q}_i^l s_i + w_k^l \quad k = 1, 2, \dots, K, \tag{4}$$

where the s_i is the pilot symbol, and \mathbf{h}_k is the complex channel vector of the k^{th} subscriber, and w_k^l is complex AWGN variable, and each element of \mathbf{h}_k and w_k^l are independent and identically distributed (i.i.d.) with zero mean and unit variance. Each subscriber knows the pilot symbol s_i well, so the SINR of the k^{th} subscriber over the m^{th} beam in the l^{th} mini-slot is:

$$SINR_{k,m}^l = \frac{|\mathbf{h}_k \mathbf{q}_m^l|^2}{1 + \sum_{i=1, i \neq m}^B |\mathbf{h}_k \mathbf{q}_i^l|^2}, \tag{5}$$

where $k = 1, 2, \dots, K$, $m = 1, 2, \dots, B$, and $l = 1, 2, \dots, L$. We denote $\gamma_m^l = |\mathbf{h}_k \mathbf{q}_m^l|^2$ as the channel gain of the k^{th} subscriber over the m^{th} beam in the l^{th} mini-slot. Then each subscriber feeds back the maximum value $SINR_{k,m}^l$ and the corresponding beam index m .

The BS computes the sum rate capacity C_{sum}^l of each $Q^l \in \mathcal{Q}$, according to the partial CSI. Here we employ the subscriber-scheduling and beam-assigning strategy of the conventional ORBF, so the sum rate capacity for the l^{th} weighting unitary matrix is estimated as

$$C_{sum}^l = \sum_{m=1}^B \log \left(1 + \max_{1 \leq k \leq K} SINR_{k,m}^l \right). \tag{6}$$

And the BS chooses the best weighting unitary matrix Q^{l^*} with the maximum sum rate capacity, that is

$$l^* = \arg \max_{1 \leq l \leq L} C_{sum}^l. \tag{7}$$

The Q^{l^*} will be performed during the time slot. So the sum rate capacity of the downlink BC channel with MWMS can be estimated as:

$$(C_{sum}^{l^*})_{MWMS} = E \left\{ \sum_{m=1}^B \log \left(1 + \max_{1 \leq k \leq K} SINR_{k,m}^{l^*} \right) \right\}. \tag{8}$$

4 Proportional Fair Scheme for MWMS

The MWMS scheme always serves those users with maximum rate in the cell, that means the channel vectors of the selected users are much more matched with the beamforming vectors. So the unfairness is introduced between users. In this section, we combine the PFS with MWMS scheme to improve the proportional fairness of the system.

The PFS for multi-user system is proposed and analyzed in Literature [9]. In this paper the PFS is used for ORBF multi-user system with multiple weight matrixes, and proposed scheme is denoted as MWMS-PFS.

In the l^{th} mini-slot, BS selects users for the m^{th} beam:

$$k_m^* = \arg \max_{1 < m < N_i} \frac{R_{k,m}(t)}{T_k(t)}, \tag{9}$$

where $R_{k,m}(t)$ means the request rate of the k^{th} user on the m^{th} beam of unitary matrix Q^l , and $T_k(t)$ means the average data rate of the k^{th} user in the past t_c time slots. The average throughput of the k^{th} user is updated as follows:

$$T_k(t+1) = \begin{cases} \left(1 - \frac{1}{t_c}\right) T_k(t) + \frac{1}{t_c} R_{k^*,m}(t), & \text{if } k \in S^l \\ \left(1 - \frac{1}{t_c}\right) T_k(t) & \text{otherwise} \end{cases} . \tag{10}$$

We define S^l to be the set of the selected users for the l^{th} mini-slot. So the sum rate of the the l^{th} unitary matrix of MWMS-PFS scheme is

$$(C_{sum}^l)_{MWMS-PFS} = \sum_{m=1}^B \log(1 + R_{k^*,m}(t)) \quad k \in S^l . \tag{11}$$

And the BS chooses the best $Q_{PFS}^{l^*}$ with the maximum sum rate capacity, that is

$$l^* = \arg \max_{1 \leq l \leq L} (C_{sum}^l)_{PFS} . \tag{12}$$

The Q^{l^*} will be performed and the selected user set S^{l^*} will be scheduled during the time slot. So the sum rate capacity of the downlink BC channel can be estimated as:

$$(C_{sum}^{l^*})_{MWMS-PFS} = E \left\{ \sum_{m=1}^B \log(1 + R_{k^*,m}^{l^*}(t)) \right\} . \tag{13}$$

5 Numerical Simulation

The proposed scheme is investigated and compared with the conventional ORBF and MOB under the system model described in previous section. The fading coefficients of the time-varying Rayleigh fading channel are i.i.d. among subscribers and for different antennae, and different time slot. The BS is equipped with $N_t = 4$ antennae, and each subscriber with $N_r = 1$ antenna. The subscribers perfectly know the CSI. All the subscribers are uniformly distributed around the BS. The plots are obtained through Monte-Carlo simulations averaged over one million realizations.

Fig. 1 shows the normalized throughput of MWMS scheme, MOB Scheme and ORBF scheme for $SNR = 20dB$. The BS generates $L = 1, 2, 3, 5, 10$ unitary matrices for each time slot in MWMS scheme. We can see that when $L=1$, MWMS retrogresses to the conventional ORBF, and when $L=2$ MWMS has the same performance with MOB and when $L=3$, the capability of the proposed MWMS scheme can improves approximately 1bps/Hz to ORBF, i.e. 10 percent. And the capability is also better than MOB scheme. With the increase of L , the normalized throughput of MWMS is improved significantly.

The normalized throughput of MWMS, MOB, ORBF is shown in Fig.2 as a function of SNR . The BS generates $L = 3$ unitary matrices for MWMS. We can see that when $SNR = -10dB$, all three schemes shows almost the same performance,

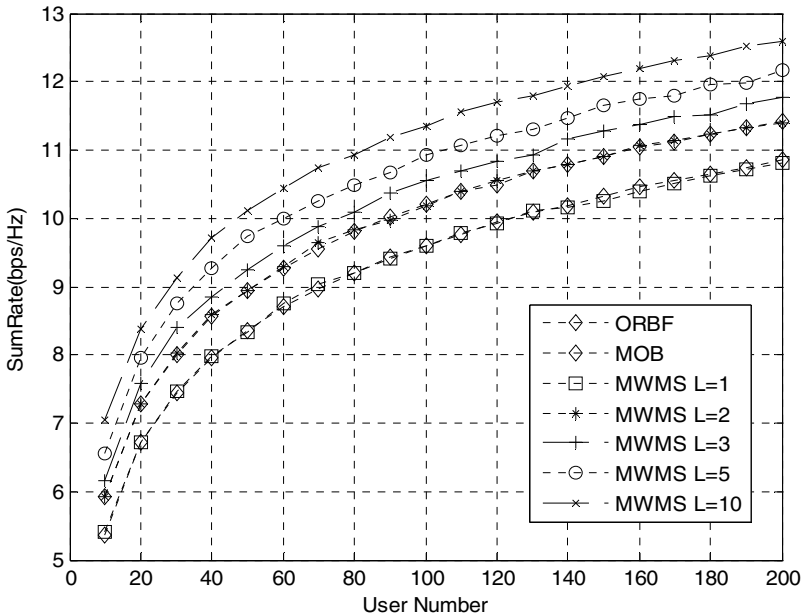


Fig. 1. Normalized throughput comparison of different number of the random unitary matrix, with $N_t = 4$, $K = 20$, $SNR = 20dB$

when $SNR = 0dB$, MWMS and MOB get about 7 percent increase from ORBF, and when $SNR = 20dB$, MWMS can supply about 1bps/Hz to the throughput. That is because when the BC channel has a low SNR , the $\gamma'_{k,m} = |\mathbf{h}_k \mathbf{q}_m^l|^2$ of each beam and each matrix is very small and it difficult to find a particular $\gamma'_{k,m}$. When the BC channel has a high SNR , it is easy to find out a matrix Q^l making the $\gamma'_{k,m}$ of each beam is much better than the others, and then the throughput increases significantly.

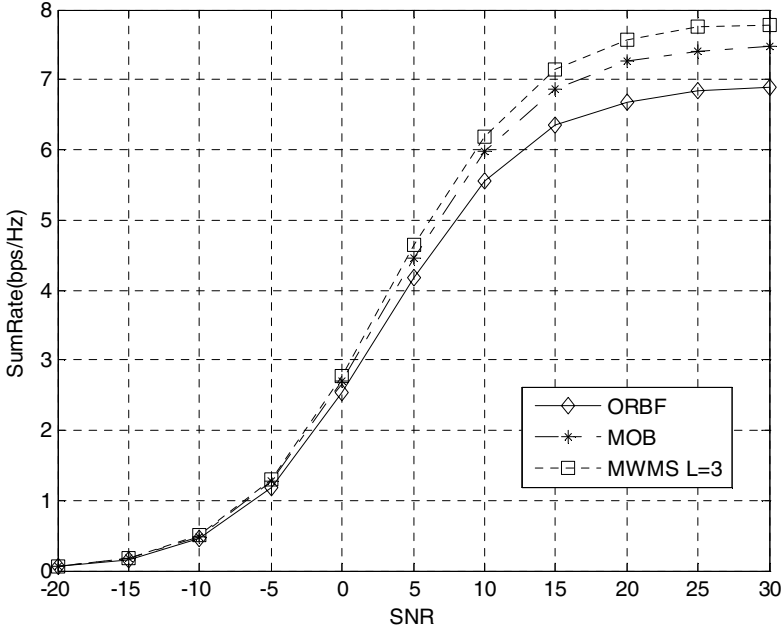


Fig. 2. Normalized throughput of MWMS, MOB, ORBF vs SNR , with $N_t = 4$, $K = 20$, $L = 3$

To evaluate the proportional fairness of the MWMS-PFS, we use a simulation scenarios with 10 users distributed normally in the cell, but the channel vectors between users are not i.i.d distribution. we assume that $\mathbf{h}_k \sim CN(0,1)$ $k = 1, 2, 3$, $\mathbf{h}_k \sim CN(0, \frac{1}{2})$ $k = 4, 5, 6$, and $\mathbf{h}_k \sim CN(0, \frac{1}{4})$ $k = 7, 8, 9, 10$, and the $t_c = 20$. We plot the average data rates of 10 users after 10000 time slots. The plot is shown as Fig.3. We can see that the MWMS-PFS scheme can increase the data rate of user 7,8,9,10, and then the fairness of MWMS-PFS is much better than the MWMS. Fig.3 also shows that the fairness of ORBF-PFS (that is the conversional ORBF with PFS scheme) is better than MWMS-PFS, but we should notice that the sum rate of the ORBF-PFS decreases much to achieve the fairness.

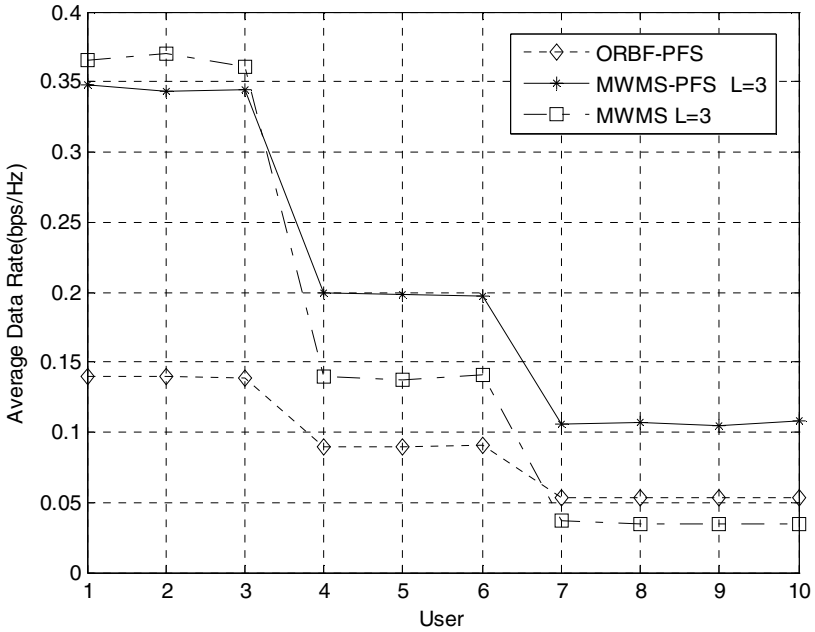


Fig. 3. Average data rate vs users of ORBF-PFS, MWMS-PFS and MWMS, with $SNR = 0$, $N_t = 4$, $K = 10$, $L = 3$

6 Conclusion

In this paper, The Downlink BC channel of multi-user MIMO system is considered. A MWMS scheme is proposed to improve the throughput of the multi-user MIMO system. The proposed scheme uses multiple random unitary matrices for the Downlink BC channel and choose a perfectly matched random unitary matrix for each time slot. when the number of subscriber in the cell is sparse the performance is improved significantly. To increase the fairness between users, we combine the MWMS scheme with the PFS, and the performance loss of sum rate is less than the ORBF-PFS.

Like ORBF, the proposed scheme needs only the partial CSI to be feed back. Because multiple random unitary matrices are used in each time slot, the feedback overhead increases as a factor of L . The efficient methods for impacting the feedback overhead over the uplink channel should be used with the proposed scheme.

Acknowledgments

This work is partly supported by International Science and Technology Cooperation Programme of China under Grant No.2008DFA11630, and No.2008AA01Z204, National Natural Science Foundation of China under Grand No.60496315 and No. 60802009, Hubei Science Foundation under Grant No.2007ABA008, and Postdoctoral Foundation under Grant No.20070410279.

References

1. Viswanath, P., Tse, D.N.C., Laroia, R.: Opportunistic Beamforming Using Dumb Antennas. *J. IEEE Transactions on Information Theory* 48, 1277–1294 (2002)
2. Caire, G., Shamai, S.: On the Achievable Throughput of A Multiantenna Gaussian Broadcast Channel. *J. IEEE Transactions on Information Theory* 49, 1691–1706 (2003)
3. Yu, W., Cioffi, J.M.: Sum Capacity of Gaussian Vector Broadcast Channels. *J. IEEE Transactions on Information Theory* 50, 1875–1892 (2004)
4. Weingarten, H., Steinberg, Y., Shamai, S.: The Capacity Region of the Gaussian MIMO Broadcast Channel. In: *IEEE International Symposium on Information Theory, Chicago*, p. 174 (2004)
5. Sharif, M., Hassibi, B.: On the Capacity of MIMO Broadcast Channels with Partial Side Information. *J. IEEE Transactions on Information Theory* 51, 506–522 (2005)
6. Kountouris, M., Gesbert, D.: Robust Multi-user Opportunistic Beamforming for sparse networks. In: *IEEE 6th Workshop on Signal Processing Advances in Wireless Communications, New York*, pp. 975–979 (2005)
7. Baran, I.R., Uchoa-Filho, B.F.: Exploiting Time Coherence in Opportunistic Beamforming for Slow Fading Channels. In: *IEEE Wireless Communications and Networking Conference, Las Vegas*, pp. 1753–1758 (2006)
8. Li, G., Yih-Fang, H.: An Opportunistic Downlink Mimo-Ofdm Scheme. In: *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 285–288 (2006)
9. Kountouris, M., Gesbert, D.: Memory-based Opportunistic Multi-user Beamforming. In: *International Symposium on Information Theory Proceedings*, pp. 1426–1430 (2005)
10. Il-Min, K., Seung-Chul, H., Ghassemzadeh, S.S., Tarokh, V.: Opportunistic Beamforming based on Multiple Weighting Vectors. *J. IEEE Transactions on Wireless Communications* 4, 2683–2687 (2005)
11. Wan, C., Forenza, A., Andrews, J.G., Heath, R.W.: Opportunistic Space-Division Multiple Access With Beam Selection. *J. IEEE Transactions on Communications* 55, 2371–2380 (2007)
12. Hassibi, B., Marzetta, T.L.: Multiple-antennas and Isotropically Random Unitary Inputs: The Received Signal Density in Closed Form. *J. IEEE Transactions on Information Theory* 48, 1473–1484 (2002)

A Novel Adaptive Reclosure Criterion for HV Transmission Lines Based on Wavelet Packet Energy Entropy

Yuanyuan Zhang, Qingwu Gong, and Xi Shi

School of Electrical Engineering, Wuhan University,
Wuhan 430072, China
zyycom@sina.com

Abstract. A novel adaptive reclosure criterion based on energy entropy to identify faults for HV transmission lines is presented in the paper on the basis of the introduction of wavelet packet and energy entropy. The method analyses the faulted phase voltage after breakers trip. After breaker trips, the voltage of transient faults is obviously different from that under permanent faults because the secondary arc has the process of extinction-reburning, and the criterion can identify fault nature before secondary arc extinction. Firstly, the faulted phase voltage is decomposed by three-layer wavelet packet, and eight signals in the third level are reconstructed. Secondly, the energy entropy under each scale is calculated according to reconstructed signals, and calculation results show the vectors under different faults are of different characteristics. Finally, a target function is defined as fault identification parameter. Simulations in EMTP/MATLAB show that: this method can determine fault nature rapidly and accurately and it won't be affected by system operation manner, transition resistance and fault location so it will have a good adaptability.

Keywords: Wavelet packet energy entropy(WPEE), HV transmission line, single-phase adaptive reclosure, Permanent faults, Transient faults.

1 Introduction

Statistics show that over 70% of faults on high-voltage transmission lines are single-phase grounded faults, of which more than 80% are transient faults. It's essential to identify the fault nature to avoid reclosure on permanent faults[1]. In recent years, many scholars have studied the single-phase adaptive reclosure and proposed many methods to distinguish between transient and permanent faults, which are typically based on voltage criterion[2,3] and harmonics detection[4,6]. The sensitivity of the former is influenced by some factors such as system operation manner, transition resistance and so on. The current research of the latter mainly concentrates on harmonic content analysis(total harmonic distortion factor e.g.[6]). It's sensitivity is easily influenced by harmonic magnitude. The secondary arc is the typical characteristic of transient faults. Wavelet packet can subtly decompose signal low-frequency components and high-frequency components simultaneously. It's fit to detect high-frequency signals produced by the secondary arc.

Based on statistical probability, information entropy is a description for the uncertainty degree of the system. In recent years it firstly begins to be applied in the biomedical, fault diagnosis, and have achieved some fruitful results[7]. But its application in power system is for not a long time, application fields and application effects are worth studying. Paper[8,9] propose the assumption of utilizing wavelet entropy in power system fault diagnosis and give several definition of wavelet entropy. A novel adaptive reclosure criterion to identify faults for HV transmission lines is presented in this paper on the basis of the introduction of wavelet packet and energy entropy basic principle. The method uses wavelet packet energy entropy(WPEE) to detect quantitatively time-frequency distribution characteristics of the faulted voltage, which can accurately identify the secondary arc so as to determine the fault nature. Furthermore, the sensitivity is not influenced by system operation manner, transition resistance and harmonic magnitude. The validity of the criterion is verified by plenty of simulations in EMTP/MATLAB.

2 Calculation Method of WPEE for Faulted Phase Voltage

2.1 Wavelet Packet Decomposition of Signal

Wavelet packet transform is a more detailed method for signal decomposition and reconstruction from the extension of wavelet analysis. The essence of wavelet analysis is the multi-scale analysis to signal S , namely:

$$V_0 = W_1 \oplus V_1 = W_1 \oplus W_2 \oplus V_2 = \bigoplus_{j=1}^J W_j \oplus V_J \quad (1)$$

Where J is decomposition scale; V_j and W_j ($j=1, 2, \dots, J$) are respectively the sub-space obtained by V_0 orthogonal decomposition under different scales. The difference between wavelet packet and wavelet analysis lies in that the former can decompose two sub-space, namely:

$$V_0 = U(1,0) \oplus U(1,1) = [U(2,0) \oplus U(2,1)] \oplus [U(2,2) \oplus U(2,3)] = \bigoplus_{b=1}^B U(J,b) \quad (2)$$

So wavelet packet has the ability to decompose signal low-frequency components and high-frequency components simultaneously. It improves the decomposition characteristic for high-frequency partial components on the basis of wavelet analysis. What's more, it can choose appropriate frequency band which matches the signal spectrum according to the signal characteristics, resulting in improved time-frequency resolution.

2.2 Definition of WPEE

Firstly sequences $S_{j,k}$ ($k=0, 1, 2, \dots, 2^j-1$) are obtained after j layer wavelet packet decomposition of signal S . Then divide $S_{j,k}$ into N pieces according to the signal time characteristics and calculate the signal energy of each piece, namely:

$$Q_i(j, k) = \int_{t_{i-1}}^{t_i} |A_i(t)|^2 dt \tag{3}$$

Where $A_i(t)$ is the i -th piece signal magnitude($i=1, 2, \dots, N$). t_{i-1} 、 t_i are the beginning and end time of the i -th piece. Normalize the signal energy of each piece $Q_i(j, k)$, then get normalized value $p_{j, k}(i)$, namely:

$$p_{j, k}(i) = \frac{Q_i(j, k)}{\sum_{i=1}^N Q_i(j, k)} \tag{4}$$

As a information measure of the system in a certain state, information entropy measures the degree of system disorder. When used in signal analysis, it can measure the signal uniformity or complexity[10]. Based on the basic theory of information entropy, the energy entropy of the k node signal of wavelet packet decomposition on j layer is defined as:

$$H_{j, k} = -\sum_{i=1}^N p_{j, k}(i) \lg p_{j, k}(i) \tag{5}$$

Define when $p=0, p \times \lg p=0$.

2.3 Extraction Steps of WPEE

As is known from the above entropy theory that energy entropy can reflect the energy distribution of the assigned frequency band. For faulted phase voltage on HV transmission line, their entropy values under different faults are different because their energy distribution characteristics are not same. The entropy contains the essential characteristics of faults. After wavelet packet decomposition, reconstruct the node signal to calculate energy entropy, so as to represent quantitatively the time-frequency energy distribution of signals with WPEE. The extraction steps of energy entropy are as follows:

- 1) Decompose original signal S to J layer with wavelet packet transform, and reconstruct signal on 2^J nodes of layer J respectively.
- 2) Then divide reconstructed node signal into N pieces on average and calculate the signal energy of each piece with (3). Normalize the signal energy of each piece $Q_i(j, k)$, then get normalized value $p_{j, k}(i)$ as shown in (4).
- 3) Calculate the energy entropy of each node signal, and utilize the entropy on layer J to form the entropy vector $T=[H_0, H_1, \dots, H_{2^J-1}]$.
- 4) Deal with the faulted phase voltage under transient fault and permanent fault according to Step 1 to Step 3, then get the corresponding entropy vector.

3 Principle of Determining Fault Nature

When faults occur on HV transmission lines, plenty of high-frequency transient components will be generated. They are relative to line parameters, fault conditions and irrespective to system operation manner, transition resistance[11]. Therefore, the fault

nature criterion based on transient components is not influenced by power frequency phenomenon(system oscillation e.g.), transition resistance and other factors.

For permanent faults, high-frequency signals produced at fault instant attenuate quickly after breakers trip, so the faulted phase voltage energy mainly lays on low-frequency components. For transient faults after breakers trip, the arc don't become extinguished immediately and will maintain rather a long time. The arc in the period is called the secondary arc. It will go through the repeated process of extinction-reburning until the arc voltage is no longer bigger than arc re-ignition voltage, then it just enters in real extinction state. The non-linear character of the arc can cause the fault phase voltage distortion, the high-frequency signals produced by it will make the time-frequency distribution characteristics of faulted phase voltage significantly different from the normal state and a permanent fault.

WPEE can represent quantitatively time-frequency distribution characteristics of the signal in appointed time period, which can accurately detect the secondary arc so as to determine the fault nature. As is known from the definition of $H_{j,k}$, it is mainly determined by signal energy time distribution under k scale and not influenced by the harmonic absolute magnitude. While the arc characteristics determine the energy distribution of faulted phase voltage, the sensitivity is not influenced by harmonics magnitude when utilizing the energy entropy to measure the fault characteristic. Compared to the approach of using voltage harmonic content alone to determine fault nature, the identification criterion with energy entropy is more accurate and reliable.

4 Simulation

4.1 Simulation System and Parameters

This paper adopts EMTP-RV to realize digital simulation experiments. The faulted phase voltage waveforms in different faults are obtained. Then they are analyzed by the wavelet tools in MATLAB. Take samples to the voltage waveform before and after breakers trip time, and the sampling frequency is 20kHz. The data window length for entropy calculation is chosen as 100ms, 20ms before trip and 80ms after trip. Transmission line system is shown in Fig.1. Where line parameters: $R_l=0.019\Omega/\text{km}$; $R_0=0.1675\Omega/\text{km}$; $L_l=0.9134\text{mH}/\text{km}$; $L_0=2.719\text{mH}/\text{km}$; $C_l=0.014\mu\text{F}/\text{km}$; $C_0=0.00834\mu\text{F}/\text{km}$.

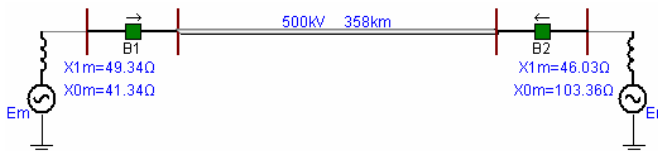


Fig. 1. 500kV HV transmission system

Assume fault occurs at 0.1s, breakers trip after two fundamental cycles and reclose at 0.6s. For single-phase transient faults, the secondary arc model is referred to paper[12,13]. Arc model parameters are as follows: $\alpha = 0.5$, $\tau_0 = 1\text{ms}$, $l_0 = 2.3\text{m}$,

$u_0 = 1100 \text{ V}$, $dr/dt = 30M\Omega/(s\sqrt{m})$, $r_0 = 22.0 \text{ m}\Omega$, $g_{\min} = 40\mu S/m$. For more general analysis, the simulated voltage waveform is analyzed under different phase-angle δ_{MN} , fault locations d , transition resistances R_g . Fig.2,3 are respectively the simulation waveforms for transient faults and permanent faults.

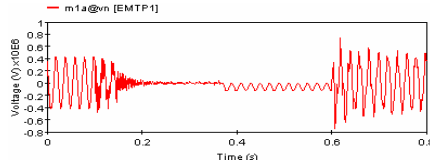


Fig. 2. Faulted phase voltage waveform for transient fault

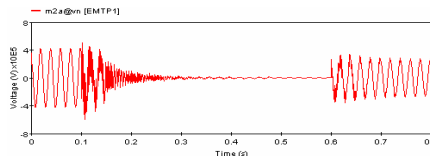


Fig. 3. Faulted phase voltage waveform for permanent fault

4.2 WPEE Calculation and Adaptive Reclosure Criterion

Wavelet base selection will affect the accuracy of signal analysis. Daubechies wavelet function series are rather sensitive to irregular signals and adaptive to analysis of transient components[14]. This paper chooses db4 wavelet for wavelet packet transform. The layer number of decomposition determines the frequency resolution.

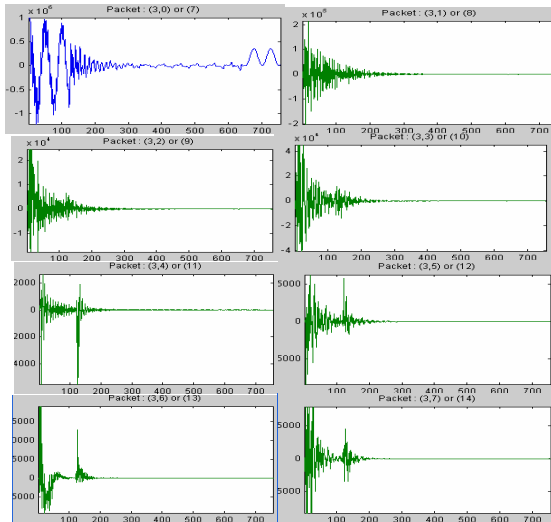


Fig. 4. Wavelet packet decomposition of faulted phase voltage waveform for transient faults

The fewer the layers are, the smaller amount of calculation, but the lower frequency resolution. While too many layers will increase calculation amount and affect the signal real-time processing. Taking into account the above factors, db4 wavelet base and layer number $J=3$ are chosen for the faulted phase voltage wavelet packet decomposition after plenty of comparative tests, and finally we get 8 node signals on layer 3.

The node signals of voltage decomposition for transient faults and permanent faults during 0.1s~0.4s are shown respectively in Fig.4 and Fig.5. The horizontal coordinate is the sampling points. As is seen from the figures that each node amplitude of the voltage attenuates quickly in a single decrease trend under the permanent fault, while the high-frequency node signal amplitude becomes larger suddenly at the moment of arc reburning and attenuates to zero through a rather long time under the transient fault.

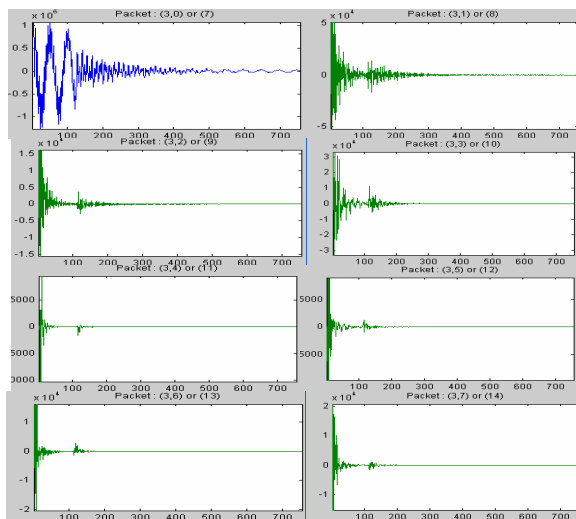


Fig. 5. Wavelet packet decomposition of faulted voltage waveform for permanent faults

After obtaining decomposition signals, set $N=100$, and thus the energy entropy values in Table 1 can be calculated according to (3)~(5). The table gives entropy calculation results of two kinds of faults with 0.12s~0.22s data in different fault conditions.

Table 1. WPEE values of faulted phase voltage

Fault conditions	Fault type	H_0	H_1	H_2	H_3	H_4	H_5	H_6	H_7	$C(H)$
$\delta_{MN}=15^\circ, R_g=0\ \Omega, d=30\%$	Permanent	1.225	1.063	0.739	0.358	0.361	0.217	0.185	0.059	0.280
	Transient	1.361	1.228	1.228	1.236	1.292	1.175	1.086	0.791	0.594
$\delta_{MN}=30^\circ, R_g=0\ \Omega, d=50\%$	Permanent	1.233	1.035	0.849	0.421	0.353	0.220	0.232	0.046	0.290
	Transient	1.353	1.261	1.263	1.244	1.286	1.169	1.062	0.843	0.591
$\delta_{MN}=15^\circ, R_g=100\ \Omega, d=80\%$	Permanent	1.218	1.027	0.732	0.329	0.370	0.215	0.134	0.051	0.270
	Transient	1.462	1.254	1.255	1.238	1.273	1.127	1.105	0.872	0.586

As is seen from the Table 1 that the entropy values of transient faults and permanent faults on low-frequency node are rather close especially for H_0, H_1 , but they have a great difference on high-frequency nodes. In theory, it could be interpreted that the existence of second arc leads to the more complexity in the voltage time-frequency distribution than the permanent fault.

To represent the entropy value difference more conveniently, define the target function:

$$C(H) = \frac{\sum_{i=3}^7 H_i}{\sum_{i=0}^7 H_i} \quad (0 < C \leq 1) \quad (6)$$

It is the ratio of the high-frequency node entropy values to all entropy values. As can be seen from Table 1, for permanent faults, $C(H)$ of the faulted phase voltage mainly lies in the range [0.27-0.3], and it lies in the range [0.58-0.6] for transient faults. Moreover, the results are the same in a large quantity of simulations. So we can get a larger margin of the threshold $C_{set} = 0.45$, and come to the criterion: $C(H) > C_{set}$. If the calculated $C(H)$ exceeds the threshold value, identifying it as a transient fault, otherwise as a permanent fault.

5 Conclusion

Based on the introduction of wavelet packet and energy entropy, this paper proposes a novel adaptive reclosure method that utilizes WPEE to describe quantitatively time-frequency distribution characteristics of the faulted voltage, and determines fault nature by entropy difference. Simulations for the transient fault and permanent fault are conducted in different conditions in EMTP. Energy entropy vector is obtained by analyzing the faulted phase voltage waveform in MATLAB. Theoretical analysis and simulation results prove that: the scheme can accurately determine the fault nature before the secondary arc extinction and not be affected by system operation manner, fault location, transition resistance. Furthermore, the sensitivity is not influenced by harmonic magnitude. Therefore, it is a effective method and of great significance for the adaptive reclosure research.

References

1. Ge, Y.Z.: New Types of Protective Relaying and Fault Location Theory and Techniques. Xi'an Jiaotong University Press (1996) (in Chinese)
2. Fan, Y., Shi, W.: Modification of Voltage Criterion in the Single-pole Automatic Reclosing of Transmission Lines. Automation of Electric Power Systems 24, 44-47 (2000) (in Chinese)
3. Li, B., Li, Y.L., Sheng, K., et al.: The Study on Single-pole Adaptive Reclosure of EHV Transmission Lines with the Shunt Reactor. Proceedings of the CSEE 24, 52-56 (2004) (in Chinese)
4. Liu, H.F., Lin, X.N., Liu, P., et al.: An Adaptive Single-phase Auto-reclosure Scheme Based on Morphological Close-opening-open-closing Gradient (COOCG) Transform. Automation of Electric Power Systems 29, 39-44 (2005) (in Chinese)

5. Li, B., Li, Y.L., Zeng, Z.A., et al.: Study on Single-pole Adaptive Reclosure Based on Analysis of Voltage Harmonic Signal. *Power System Technology* 26, 53–57 (2002) (in Chinese)
6. Radojevic, Z.M., Shin, J.R.: New Digital Algorithm for Adaptive Reclosing Based on the Calculation of the Faulted Phase Voltage Total Harmonic Distortion Factor. *IEEE Transaction On Power Delivery* 22, 37–41 (2007)
7. Gui, Z.H., Han, F.Q.: Neural Network based on Wavelet Packet-characteristic Entropy for Fault Diagnosis of Draft Tube. *Proceedings of the CSEE* 25, 99–102 (2005) (in Chinese)
8. He, Z.Y., Liu, Z.Q., Qian, Q.Q.: Study on Wavelet Entropy and Adaptability of Its Application in Power System. *Power System technology* 28, 17–21 (2004) (in Chinese)
9. He, Z.Y., Cai, Y.M., Qian, Q.Q.: A Study of Wavelet Entropy Theory and Its Application in Electric Power System Fault Detection. *Proceedings of the CSEE* 25, 38–43 (2005) (in Chinese)
10. He, Z.Y., Chen, X.L., Luo, G.M., et al.: Faulted Phase Selecting Method of Transmission Lines Based on Wavelet Entropy Weight of Transient Current. *Automation of Electric Power Systems* 30, 39–43 (2006) (in Chinese)
11. Xia, M.C., Huang, Y.Z., Wang, X.: Development and Present Situation of Transient Based Protections for High Voltage Power Transmission Lines. *Power System Technology* 26, 65–69 (2002) (in Chinese)
12. Dudurych, I.M., Gallagher, T.J., Rosolowski, E.: Arc Effect on Single-phase Reclosing Time of a UHV Power Transmission Line. *IEEE Transactions on Power Delivery* 19, 854–860 (2004)
13. Kizilcay, M., Ban, G., Prikler, L., Handl, P.: Interaction of the Secondary Arc with the Transmission System. *IEEE Bologna PowerTech Conference* 471 (2003)
14. He, Z.Y., Qian, Q.Q.: The Electric Power System Transient Signal Wavelet Analysis Method and Its Application. *Proceedings of the EPSA* 14, 1–5 (2002)

Pre-estimate on Transport Volume of Container in Xiangjiang Catchment

Jian-Lan Zhou^{1,2}

¹ Department of Work Safety, China Three Gorges Project Corporation, Yichang, Hubei, 443002, China

² Department of Control Science and Technology, Huazhong University of Science & Technology, Wuhan, Hubei, 430074, China
ZHOUJL1999@163.com

Abstract. This paper analyzes systematically common forecasting methods for transport volume of container in Xiangjiang catchment, and applies synthetically quantitative forecasting methods such as regression analysis, traffic method-sharing model, neural networks, etc. The results are proved reasonable after comparing with related data, and may serve as reference for throughout or transport volume pre-estimation of other river catchments. It also suggests that the marine conveyance container transport on Xiangjiang River take several measurements to improve the competitive power in many transport methods.

Keywords: Pre-estimation; Transport volume of container; Xiangjiang river catchment.

1 Introduction

Xiangjiang River is the maximum river in Hunan, and is also the one of the main inland channels that connects Changjiang water system and Zhujiang water system. It takes its source at Haiyang Mountain, Lingchuan County, Guangxi Province; enters into Hunan Province at Douniuling, Quanzhou; pours into Dongting Lake at Haohekou, Xiangyin through Pingdao, Lingshuitan, Hengyang, Zhuzhou, Xiangtan and Changsha; and runs into Changjiang River at Chenglingji, Yueyang through Xiangjiang flood passage of East Dongting Lake after gathering together with Zijiang River, YuanShui River and Lishui River. At present, the comprehensive transport system by the center of Changsha that five transport ways of railway, highway, marine conveyance, aviation and pipeline develops coordinately, connects other cities, provinces and world has been set up, and plays very important role in the reform and opening, and economy development.

By the end of 2003, the total highway mileage open to traffic of five cities in Xiangjiang catchment of Changsha, Zhuzhou, Xiangtan, Hengyang and Yueyang was 27888km; the inland river mileage open to traffic was 3388km, and the civil aviation route open to traffic was 39. the passenger traffic volume in 2003 was 335,400,300 persons, and the freight volume was 254,564,300t.

2 Overview

The container transport on internal feeder lines of Hunan mainly is the international container, and the goods categories mainly are import and export goods of Hunan.

The transport volume of marine conveyance container transport on Xiangjiang River had increased from 4169 TEU in 1995 to 75250 TEU in 2004 (table 1) with an annual growth rate of 189%; the transport ship types had changed from pushboats and barges to special self propelled vessels for containers; and the ports had developed from one container dock, two berths at Xihuqiao, Changsha to four ports of Zhuzhou, Xiangtan, Changsha and Chenglingji, Yueyang, eight berths.

Table 1. Hunan Marine Conveyance Container Transport Volume in 1995-2004 (Unit: TEU)

Year	1995	1996	1997	1998	1999
<i>Transport volume</i>	4169	5918	4778	5768	9412
Year	2000	2001	2002	2003	2004
<i>Transport volume</i>	13552	39539	48269	67132	95524

-- From Hunan Provincial Communications Depts.

Take the transport of 20-foot international standard container from Changshan to Shanghai by the end of 2004 for example:

Price of marine conveyance container transport: RMB 2450Yuan/TEU;

Time of marine conveyance container transport: 3-6 days.

Xiangjiang catchment throughput prediction method is divided into qualitative and quantitative forecast prediction method, commonly used in quantitative prediction time series prediction method, regression analysis, forecasting and elasticity, such as law, as some experts predict prediction method to investigate sources of law and subjective probability Law [1-3]. In transport volume of container forecast analysis should be based on the characteristics of the river itself, to explore the smallest error of prediction methods in order to obtain more reasonable results, and planning for river development to provide the right basis.

3 Pre-estimation on Transport Volume of Container

3.1 Analysis on Development Tendency of Container

The Hunan container on Xiangjiang River has been keeping a strong development tendency from 1998, and the increment is obvious in the near future (refer to table 1 and figure 1).

Based on the survey, at present, the required trains is more than 8000 in a day in Hunan, but just about 2500 trains could be satisfied, and the satisfaction rate is less than 30%. The goods and container flow that could be transported by train shall be transported by highway and waterway. Because of the loading-limitation measurement of the highway transport, the goods exported from Guangzhou and Chenzhen will be transported by water. Moreover, the completion of New Xianing Port in Changsha improves the port conditions of Hunan for waterway containers, and provides the hardware guarantee on the development of Hunan waterway container transport.

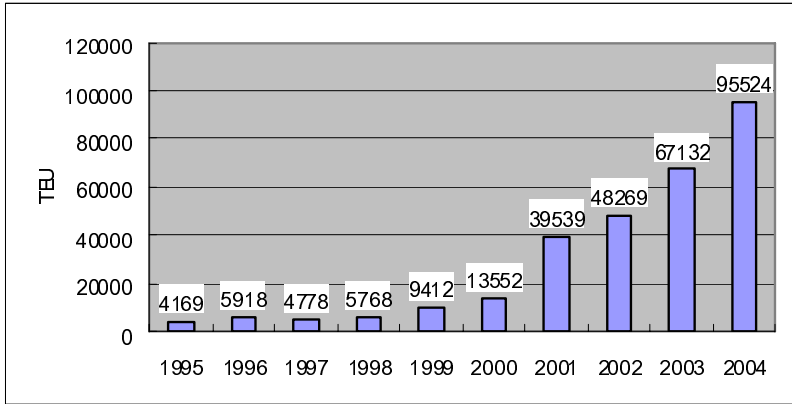


Fig. 1. Transport Volume of Container on Xiangjiang River in the Past Years

3.2 Pre-estimation Thoughts, Methods and Outcomes

According to the preestimation outcome of goods that are fit for container transport in Xiangjiang catchment, the Traffic Method-sharing Model will be used to get the transport volume of container on Xiangjiang River; the Equal-dimension Gray Step-supplement Model of Time Series, the Simple Linear Regression Model of Causality and the BP Neural Network Model of High Nonlinear could be used to preestimation.

3.2.1 Traffic Method-Sharing Model (TMSM)

The traffic method-sharing means making pre-estimation on traffic methods chosen by the consignors, making sure the transport volume undertaken by each traffic method from the given starting and ending. It is supposed, there are M zones in the planning regions, and N (N>1) transport methods in the zones. The starting and ending of the goods flow are the nodes of the zones, so the undertaking of goods transport methods could describe: there are M nodes in the transport net, and N transport methods among the nodes; the amount of Hth goods from node I to node J is X_{ijh}, and the goods volume by different methods X_{kijh} shall be confirmed. In this report, it is supposed the distribution of goods flow is known, and an improved abstract model for methods shall be used to change the good volume between starting and ending to the good volume by different methods between starting and ending. In the abstract model for methods shall not include any variable that relates to special methods. So, the Abstract Goods Transport Method-sharing Model is not only suitable for the sharing amount of existing transport methods, but also suitable for pre-estimation on sharing amount of future transport methods. So, the transport volume of container by Xiangjiang River will be gotten.

In the practice, in each transport methods, the service property could be best or not. Whether the consignor consider the important service property or not, the standards shall be the best of all the reference value. So the sharing rate of goods transport methods is the function of relative value of service properties in the transport methods.

$$P_k(X_{ijh}) = f(R_{rk}, C_{rk}, T_{rk}, F_{rk}, Q_{rk})$$

In the formula, the $P_k(X_{ijh})$ means ratio of transport volume of goods h undertaken by methods K from node i to node j ; the R_{rk} means the service dependability of transport method K , which equals to the ratio of delivery times on schedule or the consignor's appraisal; the C_{rk} means the relative transport service cost of transport method K , which equals to lowest transport service cost in the alternatives : transport service cost of transport method K ; the T_{rk} means the relative arriving time of transport method K , which equals to the shortest transport time in the alternatives : arriving time of transport method K ; the F_{rk} means the relative service frequency of transport method K , which equals to the service frequency of transport method K : maximum service frequency in the alternatives; the Q_{rk} means the transport quality of transport methods K , which equals to shortage or damage rate of goods or are evaluated by the consignors.

The improved generalized gravity model could be:

$$\sum_{k=1}^N P_k(X_{ijh}) = 1$$

$$P_k(X_{ijh}) = a_o R_{rk}^{a1} C_{rk}^{a2} T_{rk}^{a3} F_{rk}^{a4} Q_{rk}^{a5}$$

The constraint condition:

The a_1, a_2, a_3, a_4, a_5 and a_0 in the formula are the coefficient determined by the regression. According to the transport service properties comparison of Hunan (table 2), the logarithm will be gotten in both side of formula, and the formula becomes linear equation. The least square method will be used to get the parameters of the model.

Table 2. Transport Service Properties of Hunan Container in the Future

Properties	Inland river		
	2010	2015	2020
Transport cost	1/4000	1/5500	1/7500
Arriving time	1/3	1/3	1/2
Dependability	3	3	3
Service frequency	1	1	1
Transport quality	3	3	3
Properties	Railway		
	2010	2015	2020
Transport cost	1/6000	1/8000	1/10500
Arriving time	1/2	1/2	2/3
Dependability	2	2	2
Service frequency	0.05	0.08	0.1
Transport quality	2	2	2
Properties	Highway		
	2010	2015	2020
Transport cost	1/16000	1/19000	1/22000
Arriving time	1	1	1
Dependability	3	3	3
Service frequency	0.9	0.9	1
Transport quality	1	1	1

According to the transport service properties of container in the future (table 3), the sharing rate of waterway transport in 2010, 2015 and 2020 will be 49.22%, 47.67% and 45.78% respectively.

Note:

1. Transport service charge: all the transport charges that the service demanders need to pay.
2. Arriving time: all the transport time from the consignor units start the transport to the consignee units receive the goods.
3. Service dependability: the dependability of transport limit. This is a stationarity index, and in the comparison of transport methods, the 1,2 and 3 will be used to express the good and bad of transport method, the best is 3, then 2 and the worst is 1.
4. Service frequency: the operation times of vehicle, train and ship in the unit time.
5. Transport security and quality: these two indices are also stationarity indices, and in the comparison of transport methods, the 1,2 and 3 will be used to express the good and bad of transport method, the best is 3, then 2 and the worst is 1.

3.2.2 Regression Model

The demands of container transport on inland river has a strang tendency, the regression model with good increment properties could be used. According to the the actual transport volume of containers on Xiangjiang River, the second regression model of time series will be used to pre-estimate the transport volume through the comparison on practices of maths models (refer to table 3).

Table 3. Pre-estimation Value of Regression Model

(Unit: 10000 TEU)								
Year	2005	2006	2007	2008	2009	2010	2015	2020
<i>Pre-estimation value</i>	12.042	15.102	18.512	22.273	26.385	30.848	58.429	94.785

3.2.3 Radial Based Function Neural Network (RBFNN)

The historical date of National GDP (RMB 108 Yuan), Hunan GDP (RMB 108 Yuan), national import and export value (108 US dollars), Hunan foreign trade value (108 US dollars), national port goods throughput (108 t), national goods throughput on waterway (108 t), container throughput of the coastal ports (10000 TEU) will be acted as input vector, and the transport volume of container in Xiangjiang catchment could be acted as target vector, so the transport volume of container in Xiangjiang catchment in the relative year could be pre-estimated based on the development tendency of these indexes in 2010, 2015 and 2020 (refer to table 4).

Table 4. Pre-estimation on Transport Volume of Container in Xiangjiang Catchment by RBFNN

(Unit: 10000 TEU)			
Year	2010	2015	2020
<i>Pre-estimation value</i>	29.128	45.314	64.557

3.2.4 Simple Linear Regression

The transport volume of container on Xiangjiang River in these 3 years could reflect its future development tendency, and the pre-estimation value could be gotten based on the analysis of data in 2002-2004 (refer to table 5).

Table 5. Pre-estimation on Simple Linear Regression

(Unit: 10000 TEU)

Year	2010	2015	2020
<i>Pre-estimation value</i>	23.57	35.38	47.20

3.3 Pre-estimation on Container Throughput of Xiangjiang River

The container throughput will be pre-estimated according to the historical data of container throughput of Xiangjiang River. The pre-estimation outcomes and error will be given in table 6.

Table 6. Pre-estimation Value of Throughput of Container on Xiangjiang River

(Unit: 10000 TEU)

Item	2010		
	Pre-estimation part	Possible value	Development speed
<i>Throughput of Container on Xiangjiang River</i>	45-54	48	26%
Item	2015		
	Pre-estimation part	Possible value	Development speed
<i>Throughput of Container on Xiangjiang River</i>	65~80	72	8%
Item	2020		
	Pre-estimation part	Possible value	Development speed
<i>Throughput of Container on Xiangjiang River</i>	90-110	100	7%

4 Discussion

After weighting on the models mentioned above and based on the pre-estimation value mentioned above and exports' analysis, the following outcomes will be gotten (refer to table 7):

The continuous development of Hunan foreign trade, the extension of Shanghai port and ports along Changjiang River and the macro-scale operations of ships in each ocean shipping container transport companies could bring large development potential to marine conveyance container transport on Xiangjiang River. But the railway transport capability has improved, the railway container transport technology has developed, and the railway dept. pays full attention to the container transport and reinforces the marketing. So, Hunan railway container transport will become the main competition opponent to marine conveyance container transport on Xiangjiang River. Following the construction of Hunan and national expressway network and the

Table 7. Pre-estimation Value of Transport Volume of Container on Xiangjiang River

(Unit: 10000 TEU)

Item	2010		
	Pre-estimation part	Possible value	Development speed
<i>Container goods in Xiangjiang catchment</i>	55~58	57	18%
<i>Transport volume of Xiangjiang</i>	28~32	30	21%
Item	2015		
	Pre-estimation part	Possible value	Development speed
<i>Container goods in Xiangjiang catchment</i>	83~87	85	9%
<i>Transport volume of Xiangjiang</i>	40~50	45	9%
Item	2020		
	Pre-estimation part	Possible value	Development speed
<i>Container goods in Xiangjiang catchment</i>	100~106	104	4.5%
<i>Transport volume of Xiangjiang</i>	60~70	62	7%

demands increment of Chenzhen port for Hunan inland market, the Hunan railway container transport will compete with marine conveyance container transport on Xiangjiang River in a certain degree.

In regard to models comparison, it tests the reliability of the results of the model, taking $S = 10$, $F = 1 \sim 5$, the 10 years the actual data from 1985 to 1994 as input sample and expectations output, the actual data from 1995 to 2004 as a model input, through the simulation results of this 10-year variable output with corresponding real data. The results can be seen that the forecast error of the neural network model is under 10%, the prediction accuracy is higher and the result is considerable reliable.

Through the comparison of the relative error of forecasting methods, the neural network forecasting method is the best fitting and most accurate, and it is itself a identification model. So there is no need in order to establish the actual system based on mathematical forecasting models, it might dispense with the step of forecast before modeling. However, the standard BP algorithm convergences slow, vulnerable to the shortcomings of the local minimum point.

5 Conclusion

On the basis of the analysis of several commonly used prediction method, this paper forecasts the Value of Transport Volume of Container on Xiangjiang River.

Based on the conditions mentioned above, it is suggested that the marine conveyance container transport on Xiangjiang River take the following measurements to improve the competitive power in many transport methods.

1. Exploiting the favorable conditions and avoid unfavorable ones

Making full use of good operation environments, port and navigational channel resources, catching hold of opportunities of Hunan foreign trade development and industrialization distribution along the rivers, improving the auxiliary capability and service level of relative chains of marine conveyance container transport on

Xiangjiang River, compensating the disadvantages on transport time and service dependability through the liner vessels increment and improvements of punctuality rate, enlarging the transport volume and quality and accelerating the development of marine conveyance container transport on Xiangjiang River.

2. Taking multi-logistics-service strategy based on the marine conveyance container transport

Making full use of good operation environments and port resources, perfecting and improving the existing service capability and level of marine conveyance container transport on Xiangjiang River, catching hold of opportunities of international and domestic logistics quick development to open logistics service, changing the single mode of marine conveyance container transport on Xiangjiang River, developing multiply operations such as storage, conveyance and packing, infiltrating into every chain of logistics and accelerating the change of ports and marine conveyance enterprises to the 3rd party logistics.

References

1. Liu, M.W., Wang, D.Y.: Forecasting Methods for Port Throughput Capacity. *Port & Waterway Engineering* 374, 53–56 (2005)
2. Jiang, J., Wang, H.Y.: Econometric Analysis Based on the Throughput of Container and Its Main Influential Factors. *Journal of Dalian Maritime University* 2, 83–86 (2007)
3. Bates, J.M., Granger, C.W.J.: Combination of Forecasts. *Operations Research Quarterly* 20(4), 451–468 (1969)

RTKPS: A Key Pre-distribution Scheme Based on Rooted-Tree in Wireless Sensor and Actor Network

Zhicheng Dai¹, Zhi Li², Bingwen Wang¹, and Qiang Tang¹

¹ Department of Control Science and Engineering,

Huazhong University of Science and Technology, Wuhan 430074, China

² Patent Agency Center, Huazhong University of Science and Technology,
Wuhan 430074, China

dzhcheng@126.com, lizhihust@gmail.com

Abstract. Key pre-distribution in wireless sensor and actor network (WSAN) is a challenging problem because the key pre-distribution schemes in wireless sensor network (WSN) are not well-suited for the unique features and requirements. In this paper, a new scheme of key pre-distribution in WSAN (RTKPS) is presented. To achieve the distributed and integrated secure communication scheme, the paper describes a construction of the key management tree among sink, actors, and sensors by making use of the sufficient energy capabilities, larger memory, better processing and communication capabilities of actor nodes. The analysis and simulation show that the proposed scheme can effectively save memory space and evidently improve security.

Keywords: WSN, WSAN, Key Pre-distribution, Rooted-Tree, Key Ring.

1 Introduction

Wireless sensor and actor network (WSAN) derived from wireless sensor network (WSN) refers to a group of sensors and actors linked by wireless medium to perform distributed sensing and actuation tasks. In the network, sensors gather information about the physical world, while actors coordinate and make decisions to perform appropriate actions upon the environment, which allows remote, automated interaction with the environment. The WSAN has wide applications in both civil and military fields such as microclimate control in buildings, home automation and environment monitoring, biological and chemical attack detection and battlefield surveillance [1-2].

One of the fundamental problems in sensor network security is how to bootstrap secure communications, i.e., how to establish pairwise keys between neighboring nodes. To address the bootstrapping problem in WSN, much work has been conducted in recent years [3-10]. Among them, the basic random key pre-distribution scheme (named as E-G scheme) is proposed by Eschenauer and Gligor that relies on probabilistic key shared among the sensor nodes and uses simple protocols for shared key discovery and path key establishment. The scheme is briefly described as follows. A random large pool of keys is selected from the key space. Each sensor node receives a random subset of keys (named as key ring) from the key pool before deployment. Any two nodes that are able to find the key in common within their respective key ring can use that key as their shared secret to initiate communication.

Although WSN is derived from WSN, the key pre-distribution schemes which have been proposed for WSN may not be well-suited for the unique features and requirements of WSN. The nodes in WSN which includes sensors and actors are heterogeneous. The actors are resource rich nodes equipped with better processing and communication capabilities, and they are not considered in the key pre-distribution schemes in WSN. Therefore, it is necessary to explore the novel key pre-distribution scheme in WSN.

In this paper, a key pre-distribution scheme based on rooted-tree is proposed in WSN, which constructs the key management tree among sink, actor nodes and sensor nodes by making use of sufficient energy capabilities, larger memory, better processing and communication capabilities of the actor nodes. The remainder of the paper is organized as follows. Section 2 introduces RTKPS and the construction of the key management tree. Section 3 makes a qualitative and quantitative analysis of RTKPS. A conclusion is drawn in Section 4.

2 RTKPS Scheme

In this section, we present the basic features of RTKPS, deferring its analysis and simulation for the next section.

2.1 Key Pre-distribution

Just as E-G scheme, the distribution in RTKPS consists of three phases, namely key pre-distribution, shared-key discovery, and path-key establishment.

1) The key pre-distribution phase

The key pre-distribution phase of RTKPS consists of five off-line steps as follows:

Step 1: The sink generates a large pool s of $|s|$ keys and their key identifiers, which are saved into the memory.

$$\begin{aligned}
 S &= K \cup D \\
 K &= \{k_1, k_2, \dots, k_n\}, D = \{ID_1, ID_2, \dots, ID_n\}
 \end{aligned}
 \tag{1}$$

Where k_i is one key of the pool s , while ID_i is the corresponding identifier.

Step 2: $|s_A|$ keys are drawn out of $|s|$ randomly without replacement to establish the key ring s_A of the neighboring actor A of the sink. Then the key ring s_A is loaded into the actor A .

$$\begin{aligned}
 S_A &= K_A \cup D_A \\
 K_A &= \{k_{A1}, k_{A2}, \dots, k_{A|S_A|}\}, k_{Ai} \in K \\
 D_A &= \{ID_{A1}, ID_{A2}, \dots, ID_{A|S_A|}\}, ID_{Ai} \in D
 \end{aligned}
 \tag{2}$$

Where ID_{Ai} denotes the corresponding identifier of k_{Ai} .

Step 3: $|S_B|$ keys are drawn out of $|S_A|$ randomly without replacement to establish the key ring S_B of the neighboring actor B of the actor A . Then the key ring S_B is loaded into the actor B .

Step 4: $|N_C|$ keys are random drawn out of $|S_B|$ randomly without replacement to establish the key ring S_N of the neighboring sensor N of the actor B . Then the key ring S_N is loaded into the sensor N .

Step 5: Repeat step 3 and step 4, and finally the key management tree is constructed. The model of the key management tree is described as follows: WSAW is controlled and managed by the sink as the root, the whole network is separated into various subnets; each subnet consists of one actor and some sensors, that is to say, each sensor belongs to one subnet.

2) The shared-key discovery phase

The shared-key discovery phase takes place during WSAW initialization in the operational environment, when every node discovers its neighbors in wireless communication range with which it shares keys. The actor P can discover shared-key with the actor Q of the higher subnet because the key ring of the actor P is randomly drawn out of the actor Q key ring. Likewise, the actor P can discover shared-key with the actor R of the lower subnet, while the actor P can discover shared-key with the sensor N of the same subnet.

3) The path-key establishment phase

The path-key establishment phase assigns a path-key to selected pairs of nodes in wireless communication range that do not share a key but are connected by two or more links at the end of the shared-key discovery phase. The sink, actors and sensors can establish secure communication by using shared-key. Then the key management tree that is constructed as shown in Fig. 1, in which sink is the root, actors are the branches and sensors are the leaves.

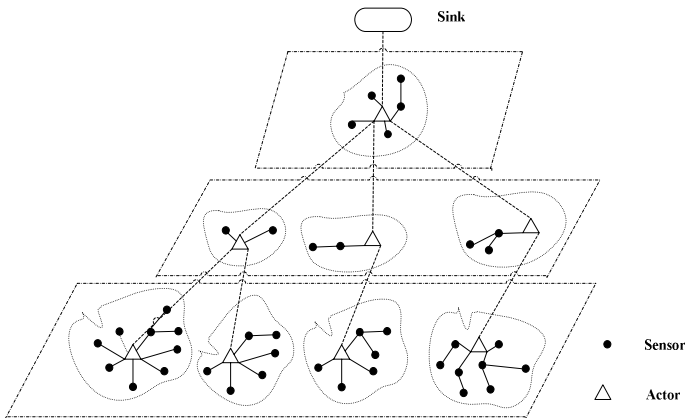


Fig. 1. The key management tree

2.2 Construction of the Key Management Tree

In RTKPS, the most important is how to construct the key management tree in the key pre-distribution phase. To describe the construction clearly, some special symbols are defined in table 1.

Table 1. The special symbols describing the construction of the key management tree

Symbol	Description
Sink	The sink node
A_i	Actor node
N_i	Sensor node
T_{th}	Time threshold, the maximal response time
k_i	Key
ID_i	Identifier of the key k_i
$M :: ()$	Message information
ACK	Acknowledge information
λ^i	Actor node belongs to i^{th} subnet
$K_{A,B}$	Shared-key of node A and B
$E(k, \dots)$	Symmetric encryption function with key k
$Broadcast :: (hello)$	Broadcast hello information
$Info$	Information including resource using condition of sink, actors and sensors

The construction algorithm of the key management tree is described as follows:

Step 1: Sink sends broadcast information. The responding actor A_i in neighborhood sends $Info$ to Sink within the given T_{th} . A key k_i that is randomly took out from the key ring of the actor A_i is used as the path key between Sink and the actor A_i (here exists shared-key between A_i and Sink because the key ring of A_i is drawn out of the key pool of Sink). Then the first subnet λ^1 is formed.

$$\begin{aligned}
 & Sink \rightarrow A_i : Broadcast(hello) \\
 & A_i \rightarrow Sink : ID_{A_i} \cup E(K_{Sink, A_i}, Info \cup M :: (t < T_{th})) \\
 & Sink \rightarrow A_i : E(K_{Sink, A_i}, M :: (ACK) \cup \lambda^1)
 \end{aligned}
 \tag{3}$$

Step 2: The actor A_{λ^i} (not sensor) in i^{th} subnet sends broadcast information. The responding nodes in neighborhood are recorded within the given T_{th} . If the responding node in neighborhood is actor A_j (not A_i), it will be the actor in the lower subnet (λ^{i+1}), while if the responding node in neighborhood is sensor N_j (not N_{λ^k} in

$\lambda^k, k \leq i$), it will be the sensor in the same subnet λ^i . If one node sends *ACK* to more than one actor within the given T_{th} , the actor that receives the *ACK* earliest is selected to be the responding node in neighborhood.

$$\begin{aligned}
 & A_{\lambda^i} \rightarrow A_{\lambda^{i+1}} : \text{Broadcast}(\text{hello}) \\
 & A_{\lambda^{i+1}} \rightarrow A_{\lambda^i} : ID_{A_{\lambda^{i+1}}} \cup E(K_{A_{\lambda^{i+1}}, A_{\lambda^i}}, \text{Info} \cup M :: (t < T_{th})) \\
 & A_{\lambda^i} \rightarrow A_{\lambda^{i+1}} : E(K_{A_{\lambda^{i+1}}, A_{\lambda^i}}, M :: (ACK) \cup \lambda^{i+1}) \\
 & A_{\lambda^i} \rightarrow N_j : \text{Broadcast}(\text{hello}) \\
 & N_j \rightarrow A_{\lambda^i} : ID_{N_j} \cup E(K_{N_j, A_{\lambda^i}}, \text{Info} \cup M :: (t < T_{th})) \\
 & A_{\lambda^i} \rightarrow N_j : E(K_{N_j, A_{\lambda^i}}, M :: (ACK) \cup \lambda^i)
 \end{aligned} \tag{4}$$

Step 3: Repeat step 2. The nodes in WSN are added to the key management tree step by step.

Step 4: After that, all actors are added to the key management tree for the actors' sufficient communication capabilities. If some sensors still have not been added to the tree, scope expanding process required is described as follows: the sensor sends broadcast information by single-hop or multi-hop paths, which is added to the subnet that the earliest responding node belongs to. The scope is expanded gradually until the network is secure complete connected graph.

2.3 Add and Delete Sensors

Usually, WSN has been arranged in the environment which is absent of duty. Sometimes the environment of the application is very bad, such as the situation of fire fighting, military reconnaissance and earthquake salvation. Sometimes, the nodes which use up the battery or be captured are deleted from the network. Sometimes, it is need to add a new node to the network. All these require WSN with the ability to self-organize and adaptive.

When a new sensor N_i is added to the network, N_i sends broadcast information to neighboring nodes. Then the earliest responding node N_0 will establish communication with N_i (no key certification).

When a sensor N_j is deleted from the network, the key ring of the subnet to which N_j belongs is required to alter. Firstly, the actor A_i of the subnet broadcasts a revocation message to all sensors of the subnet, then gains new key ring from the higher subnet. Afterwards, A_i distributes its key ring to other nodes of the subnet. If the lower subnet of A_i exists, the key ring is allocated step by step to the nodes of lower subnets. Finally, the key management tree is updated. Once N_j is removed from the

network, some links may disappear, and the affected nodes need to reconfigure those links by restarting the shared-key discovery phase and, possibly path-key establishment for them.

3 Analysis and Simulation

In WSN, the key management tree is constructed based on key pre-distribution, which ensure that all sensors and actors are established secure communication in the network initialization phase. Only there is the shared-key between the key rings of two nodes, the secure communication of the two nodes can be established. Therefore the possibility of attack is reduced and the security is improved obviously.

In RTKP, the key management tree is constructed where sink is the root, actors are the branches and sensors are the leaves, to achieve the distributed and integrated key management. The network is completely connected without information islands. If a node is added to or deleted from the network, only the key ring of some nodes is required to update and other nodes are not affected. So the security is improved and the network communication overhead is reduced.

In the subnet of actor A_i including one actor and many sensors, let P be the probability that a shared key exists between two nodes, $|N_i|$ be the number of sensors, d be the expected degree of a node. The expression in relation to P , $|N_i|$ and d is given as follows:

$$d = \left(\frac{|N_i|-1}{|N_i|} \right) (\ln(|N_i|) - \ln(-\ln(P))) \tag{5}$$

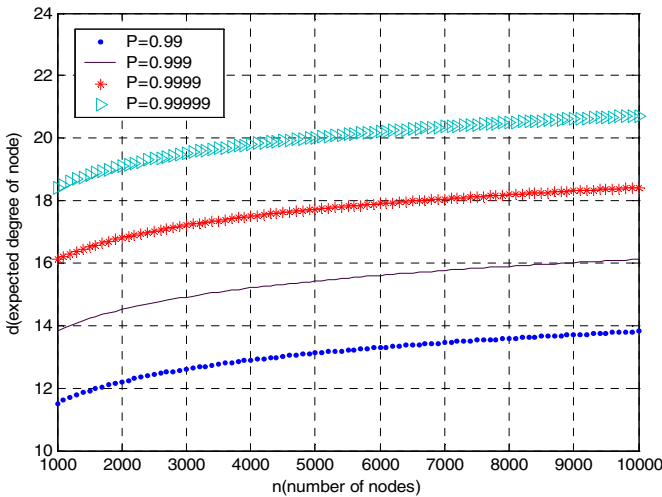


Fig. 2. Expected degree of node vs. number of nodes

Fig.2 illustrates the plot of the expected degree of a node, d , as a function of the number $|N_i|$ of the sensors, for various values of P . It shows that, to increase the probability that a random graph is connected by one order, the expected degree of a node increases too. When $|N_i|$ is large, the expected degree has almost no change, which indicates that the size of the subnet has insignificant impact on the expected degree of a node required to have a connected graph.

Let $|s|$ be the pool size, m_1 be the number of the actors' key ring in first subnet. The probability that two key rings do not share any key is given by

$$P_1 = \frac{C_{|s|}^{2m_1} C_{2m_1}^{m_1}}{C_{|s|}^{m_1} C_{|s|}^{m_1}} = \frac{[(|s|-m_1)!]^2}{(|s|-2m_1)!|s|!} \tag{6}$$

Since $|s|$ is very large, use Stirling's approximation for $n!$.

$$n! \approx \sqrt{2\pi n} n^{n+0.5} e^{-n} \tag{7}$$

So simplify the expression (6), and obtain:

$$P_1 \approx \frac{\left(1 - \frac{m_1}{|s|}\right)^{2(|s|-m_1+0.5)}}{\left(1 - \frac{2m_1}{|s|}\right)^{(|s|-2m_1+0.5)}} \tag{8}$$

The probability of actors in first subset sharing at least one key is described as $P'_1 = 1 - P_1$. To satisfy the relation $P'_1 \geq P_A$ (where P_A is the probability threshold), the relevant key ring size m_1 is required to be distributed from $|s|$ randomly. Similarly, to satisfy the relation $P'_n \geq P_A$, the relevant key ring size m_n is required to be distributed from m_{n-1} randomly.

$$P'_n = 1 - \frac{C_{m_{n-1}}^{2m_n} C_{2m_n}^{m_n}}{C_{m_{n-1}}^{m_n} C_{m_{n-1}}^{m_n}} = 1 - \frac{[(m_{n-1}-m_n)!]^2}{(m_{n-1}-2m_n)!m_{n-1}!} \approx 1 - \frac{\left(1 - \frac{m_n}{m_{n-1}}\right)^{2(m_{n-1}-m_n+0.5)}}{\left(1 - \frac{2m_n}{m_{n-1}}\right)^{(m_{n-1}-2m_n+0.5)}} \tag{9}$$

Fig.3 illustrates a plot of the connected probability P'_1 as a function of the key ring size m_1 for various $|s|$ ($=1000, 5000, 10000, 20000$). This figure shows that, when the probability threshold P_A is fixed, the larger the key pool size $|s|$, the key ring size m_1 larger is. $P'_1 = 0.5$ if $|s|=1000$ and $m_1 = 25$. To achieve the connected probability $P'_1 = 0.5$, $m_1 = 120$ when $|s|=2000$.

Fig.4 shows the probability of nodes in 1st, 2nd, 3rd, 4th subset sharing at least one key and the relevant key ring size when $|s|=50000$. It shows that, to satisfy the relation

$P_A = 0.99$, the key ring size of 1st, 2nd, 3rd, 4th subset must be at least 150, 145, 140, 135 respectively. That is to say, the key ring size decreases with the rising of subnet level because m_n is acquired from m_{n-1} . Therefore, in RTKPS, memory space of nodes can be saved, while the key ring size of all subsets is equal to that in E-G scheme.

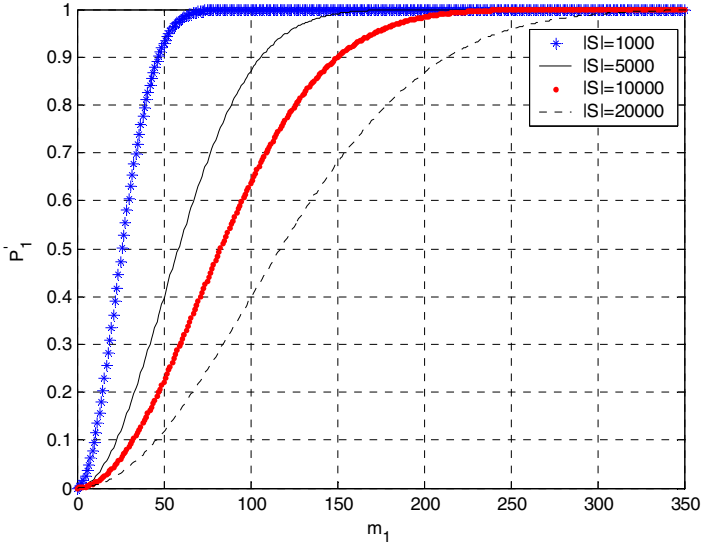


Fig. 3. Probability of sharing at least one key vs. the key ring size m_1

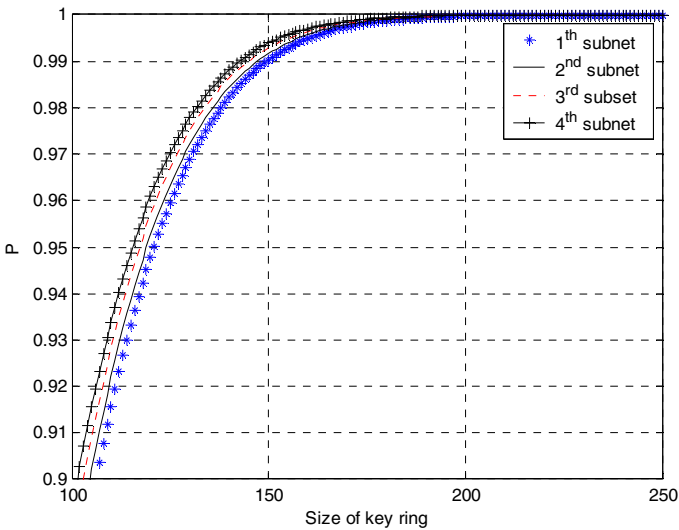


Fig. 4. Probability of sharing at least one key vs. the key ring size of 1st, 2nd, 3rd, 4th subnet

4 Conclusion

In WSN, by using key pre-distribution and dynamic distribution, the key management tree is constructed to achieve internal security communication, which is a new model presented to network security. The key pre-distribution schemes in WSN are not well-suited for the unique features and requirements. RTKPS is scalable and flexible by making use of sufficient energy capabilities, larger memory, better processing and communication capabilities of the actor nodes. The analysis and simulation indicate that our scheme is superior to E-G scheme for it saves memory space of nodes.

References

1. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., et al.: A Survey on Sensor Networks. *IEEE Communications of the ACM* 43, 51–58 (2000)
2. Akyildiz, I.F., Kasimoglu, I.H.K.: *Wireless Sensor and Actor Networks: Research Challenges*. *Ad Hoc Networks* 2, 351–367 (2004)
3. Eschenauer, L., Gligor, V.D.: A key-management Scheme for Distributed Sensor Networks. In: *Proceedings of the 9th ACM Conference on Computer and Communications Security: Association for Computing Machinery, Washington, DC, United States*, pp. 41–47 (2002)
4. Chan, H., Perring, A., Song, D.: Random Key Pre-distribution Schemes for Sensor Networks. In: *Proceedings of the IEEE Computer Society Symposium on Research in Security and Privacy*, pp. 197–213. IEEE Computer Society, Berkeley (2003)
5. Kuo, C.F., Lu, Y.F., Pang, A.C., et al.: Security Enhanced Data Delivery in Sensor Networks. In: *Proceedings of the IEEE 39th International Carnahan Conference on Security Technology*, pp. 256–259. IEEE Computer Society, Las Palmas (2005)
6. Arjan, D., Vijay, B., Vamsi, P., et al.: Secure Continuity for Sensor Networks. In: *Proceedings of 8th International Symposium on Parallel Architectures, Algorithms and Networks*, pp. 323–326. IEEE Computer Society, Los Alamitos (2005)
7. Liu, D., Ning, P., Li, R.: Establishing Pairwise Keys in Distributed Sensor Networks. In: *Proceedings of the ACM Conference on Computer and Communications Security, Association for Computing Machinery, Washington, DC, United States*, pp. 52–61 (2003)
8. Eltoweissy, M., Moharrum, M., Mukkamala, R.: Dynamic Key Management in Sensor Networks. *IEEE Communications Magazine* 44, 122–130 (2006)
9. Huang, D., Mehta, M., Medhi, D., Harn, L.: Location-Aware Key Management Scheme for Wireless Sensor Networks. In: *Proceedings of the 2nd ACM Workshop on Security of Ad Hoc and Sensor Networks*, pp. 29–42. ACM Press, New York (2004)
10. Du, W., Deng, J., Han, Y.S., et al.: A Key Management Scheme for Wireless Sensor Networks Using Deployment Knowledge. In: *Proceedings of the IEEE INFOCOM*, pp. 586–597. IEEE Press, Piscataway (2004)

Urban Road Network Modeling and Real-Time Prediction Based on Householder Transformation and Adjacent Vector

Shuo Deng, Jianming Hu, Yin Wang, and Yi Zhang

Department of Automation and TNList, Tsinghua University, Beijing 100084, China
dengshuoamanda@gmail.com, hujm@mail.tsinghua.edu.cn,
yin-wang04@mails.tsinghua.edu.cn, zhyi@mail.tsinghua.edu.cn

Abstract. This paper put forward a multivariate one-order-regression single road link model based on the algorithm of Householder Transformation to reduce the computation complexity in real-time prediction and to facilitate the study on network turn-ratio pattern evolution. Then the paper analyses the limitation of current urban road network model based on adjacent matrix and contributed a novel model based on new memory strategy aiming at reduce the memory space occupied by adjacent matrix, carrying turn movement information in the storage and avoiding redundant calculation. To verify the new modeling method, the study involved in a field work on part of urban network in Beijing, China. In conclusion, the new modeling methods in this paper enhanced the performance of urban road modeling.

Keywords: Urban road network, Householder Transformation, Adjacent Vector.

1 Introduction

At present, in the field of ITS (Intelligent Transportation Systems), the modeling of roads and the prediction of traffic flows mainly focused on single link objects. Presently, the link modeling approaches and prediction methods all derived from widely used models in other fields[1], such as auto-regression model (AR)[2], moving-average model (MA), auto-regression moving average model (ARMA)[3], history-average model (HA) [4], Box-Cox Method [5], multivariate regression model, ARIMA model [6], Kalman Filter Model[7] as well as the combination prediction models from all the above models. Besides, there are a series of non-modeling algorithms including non-parametric regression, KARIMA algorithm, wavelet network, multi-dimensional fractal based approach and various neural network related methods [8]. Of all the non-modeling algorithms, only the input and output are observable. That is inconvenient for further study on pattern recognition of the single-link [9]. While for the modeling based algorithms, the model's coefficients have little physical significance, and some of these methods implement the prediction based only on the historical data of the to-be-predicted link itself, seldom take the related links into account. So this paper brought out a new prediction modeling method: building the relationship between to-be-predicted link and its adjacent links by estimating the turn ratio. This model emphasized the dynamic characteristics of the transportation flow.

This paper is organized as follows. In Section 2, this paper first discussed the single-link modeling and flow prediction method. Section 3 focused on the structure of the entire network modeling. The simulation and verification of all the models and methods in this paper lied in Section 4, using the simulation tools TransModeler.

2 Single-Link Modeling and Prediction

2.1 Algorithm Description

In a certain intersection, the downstream flow discrete-time series $y(t)$ is directly influenced by the go-down-straight ($x_1(t)$), turn-left ($x_2(t)$) and turn-right ($x_3(t)$) flows from the upstream links. There exists a multivariate linear equation:

$$y(t+1)=a_1x_1(t)+a_2x_2(t)+a_3x_3(t) \tag{1}$$

Here, a_1 represents the ratio from go-down-straight flow’s upstream link to the downstream link. a_2 and a_3 represent the other two upstream links’ left-turn and right-turn ratio. Thus the equation describes the relation between the upstream and the downstream.

The realistic traffic flow data presents randomness; and the numerical value fluctuates greatly. Though, on the whole, the value follows the traffic habit of local population. Take the city of Beijing for example. In weekdays, the morning and afternoon traffic peak hours occur around 8:00 am and 5:00 pm. Since the working zones and living zones location differ, the former in the center of the city and the latter in the peripheral areas, the two flow peak hours’ directions differ from one another. The change in turn ratios reflects the difference. Thus, in order to build the model more practical, a_1 , a_2 and a_3 should be time-varying coefficients.

Under this model, the traffic flow’s real-time prediction can be a problem of parameter estimation of varying-time system. There are three unknown parameters. Accordingly three groups of practical data are required for initialization. Detailed idea of the algorithm presented as follow:

Choose the first three groups from the original data. Here we get the matrix of the set of parameter equations and get the solution of the initial a_1 , a_2 and a_3 :

$$\begin{bmatrix} x_1(1) & x_2(1) & x_3(1) \\ x_1(2) & x_2(2) & x_3(2) \\ x_1(3) & x_2(3) & x_3(3) \end{bmatrix} \bullet \begin{bmatrix} a_1(1) \\ a_2(1) \\ a_3(1) \end{bmatrix} = \begin{bmatrix} y(2) \\ y(3) \\ y(4) \end{bmatrix} \tag{2}$$

In the next point of time, we get another group of data from the sensor. Add it to the original data set; and then we get a set of contradictory equations:

$$\begin{bmatrix} x_1(1) & x_2(1) & x_3(1) \\ x_1(2) & x_2(2) & x_3(2) \\ x_1(3) & x_2(3) & x_3(3) \\ x_1(4) & x_2(4) & x_3(4) \end{bmatrix} \bullet \begin{bmatrix} a_1(2) \\ a_2(2) \\ a_3(2) \end{bmatrix} = \begin{bmatrix} y(2) \\ y(3) \\ y(4) \\ y(5) \end{bmatrix} \tag{3}$$

Since new group of data comes into the system sequentially, the structure of the equation set will change. In order to simplify the computation, here we adopt the recurrence least square method to deal with the added-in data to avoid re-constructing the equations every time and to reduce the load of computation [10].

In evaluation analysis, there are many approaches to realize the recurrence least square method, Householder Transformation [11] is one of the most widely used one, Liu put forward an algorithm based on Householder transformation [12]. The adapted algorithm is presented as follow:

Let $\theta(t)=[a_1x_1(t) a_2x_2(t) a_3x_3(t)]$ and $\varphi(t)=[x_1(t) x_2(t) x_3(t)]$, then the objective function for parameter estimating at time t can be defined as:

$$J(\theta(t))=\lambda J(\theta(t-1))+|y(t)-\varphi(t)^T\theta(t)|, 0<\lambda\leq 1 \tag{4}$$

Here, λ is called the ‘forget determiner’, it is used to decrease historical data’s influence over the current prediction. The smaller the λ is, the faster the system ‘forgets’ those historical data which may have subordinate influence over current prediction.

The corresponding equation set of equation (5) is:

$$\Lambda_t D_t x = 0 \tag{5}$$

Here

$$\left\{ \begin{array}{l} D_t = \begin{bmatrix} \varphi(1) \\ \varphi(2) \\ \vdots \\ \varphi(t) \end{bmatrix} \\ \Lambda_t = \text{diag}(\alpha \ \alpha \ \dots \ \alpha \ 1)_{t \times t} \\ x^T = (\theta(t)^T \ | \ 1) \\ \alpha = \sqrt{\lambda} \end{array} \right. \tag{6}$$

Construct a Householder Transformation Matrix H_t , make

$$H_t D_t = \left[\begin{array}{c|c} R_t & z_t \\ \hline 0 & g_t \end{array} \right] \tag{7}$$

Here, $R_t \in R^{t \times t}$ is an upper triangular matrix, $z_t \in R^{t \times 1}$ is a column vector.

Put

$$D_{t+1}^* = \left[\begin{array}{c|c} \alpha R_t & \alpha z_t \\ \hline \varphi(t)^T & -y(t+1) \end{array} \right] \tag{8}$$

Literature [12] proved that (5) has the same solution with $D_{t+1}^* x = 0$, that means, recurrence least square method with varying-time parameters can be realized by Householder Transformation. In practical, the realization of Householder Transformation is complex. However, considering that R_t is an upper triangular matrix, the algorithm can be further simplified.

2.3 Further Discussion on the Meaning of Forget Determiner λ

In this model, λ determines how fast the system ‘forgets’ the previous data. When $\lambda < 1$, the previous data would cast less and less influence over the prediction as time going. Correspondingly, the new data dedicates more in prediction and the system presents better performance on following features. For urban road system, the interval of data acquisition is 5 minutes. In this relatively short time-step, the possibility of data’s great change is low. Consequently, the prediction based on better following-feature system can be accurate and the estimated parameter of turn ratio can match the practical values.

Another goal of structuring this model is to study different pattern for different links focusing on the geography-oriented discrepancy. Under this goal, all the data should be treated equally. Hence, $\lambda=1$, the algorithm degenerates to basic recurrence least square method without ‘forget determiner’. In Chapter 4, simulation and analysis show the pattern.

3 Traffic Network Modeling

3.1 Adjacency Matrix Model

An Adjacency Matrix Model was introduced to describe the adjacent relationship between different links [13]. Adapting this model for the urban road network, the specific steps are listed as follows:

For all the links, construct an adjacent matrix A . Each row presents the adjacent relationship with certain links in the network. If link i is the downstream link of link j , assign $A(i,j)$ a non-zero value, the value is the turn ratio. Then we assign ‘0’ to all the other elements. With the varying-time parameters, for the entire road network, we can construct a set of equations as follows: (n represents the total number of the links in the network.)

$$\begin{bmatrix} x_1(1) & x_2(1) & x_3(1) & \cdots & x_n(1) \\ x_1(2) & x_2(2) & x_3(2) & \cdots & x_n(2) \\ x_1(3) & x_2(3) & x_3(3) & \cdots & x_n(3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_1(t) & x_2(t) & x_3(t) & \cdots & x_n(t) \end{bmatrix} \cdot A = \begin{bmatrix} y_1(2) & y_2(2) & y_3(2) & \cdots & y_n(2) \\ y_1(3) & y_2(3) & y_3(3) & \cdots & y_n(3) \\ y_1(4) & y_2(4) & y_3(4) & \cdots & y_n(4) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_1(t+1) & y_2(t+1) & y_3(t+1) & \cdots & y_n(t+1) \end{bmatrix} \tag{9}$$

Solve this set of contradictory equations with least square method then get the turning ratios for every link. Thus, the non-zero values in the adjacent matrix get modified by the influence of newly coming data. Left multiply the newest group of flow data from all the links with A then we get the prediction of the flow values of the entire network.

3.2 Adjacent Vector Modeling

Urban road network includes large amount of links, often at the magnitude of 10^3 . Put the total number of the roads as $N \times 10^3$. For most of the intersections which are four approaches, every downstream link has three direct upstream links. So the adjacent matrix is a sparse matrix which causes severe waste of memory space. Besides, the adjacent matrix only expresses the logic adjacent relation, without the turning direction information. Moreover, adopting the left multiplying algorithm will cause many times of ‘multiply by zero’ calculation, which decreases the computation efficiency. Therefore, we introduce a new structure as follows:

Express the adjacent relation with ‘Adjacent Vector’:

$$v = \begin{pmatrix} \text{downstream_link-ID} \\ \text{upstream_go_straight_link-ID} \\ \text{go_straight_ratio} \\ \text{upstream_turn_left_link-ID} \\ \text{turn_right_ratio} \\ \text{upstream_turn_right_link-ID} \\ \text{turn_left_ratio} \end{pmatrix} \tag{10}$$

When predicting next time-step’s flow, multiply the ratios with current flow data with corresponding link ID.

The advantages of the ‘Adjacent Vector’ Model lie in:

- Compared with adjacent matrix, this model occupies smaller memory space with the scale of $7 \times N \times 10^3$.
- This model carries the information of the geography relationship between upstream and downstream links, including go-straight, turn-left and turn-right.
- The prediction only focuses on the non-zero sub-blocks, avoiding redundant ‘multiply zero’ calculation, enhancing the computation efficiency.

4 Simulation and Verification

4.1 Construction of Road Network

We choose a realistic network inside the 2nd Ring Road of Beijing, China, including 81 intersections and 118 main links. Using the simulation software TransModeler, we introduce field map into the simulation project. Besides, we carry an investigation of the numbers of link, traffic signal time-distributing. Therefore, the proposed model shares a high level of consistency with the realistic road network.

4.2 Verification

Predict the flow by the Householder Transformation based least square method, the prediction result is shown in Figure 1.

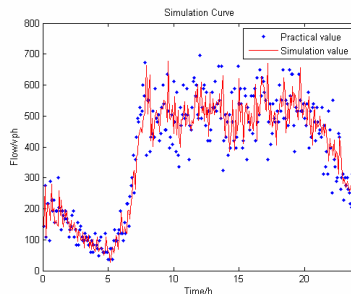


Fig. 1. The average relative error of the simulation is 20.17%, an acceptable error in transportation flow prediction. ($\lambda=0.79$)

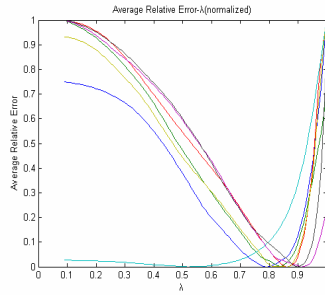


Fig. 2. Average relative errors vary while λ changes in different road links

In Figure 2, When λ approaches zero—means prediction only based greatly on new data—the prediction bears great error because of random fluctuation character of practical traffic data. On the other hand, when λ approaches 1—means prediction based equally on new data and previous data—the prediction also bears great error since historical data serves as interference. For different links, the optimal λ values (with minimal average relative error) concentrate in a narrow scale of (0.75,0.95). This scale represents the stochastic characteristic of transportation flow. Thus, in order to get more accurate prediction, the λ should be assigned in this scale.

The estimation of the turn ratios is shown in Figure 3.

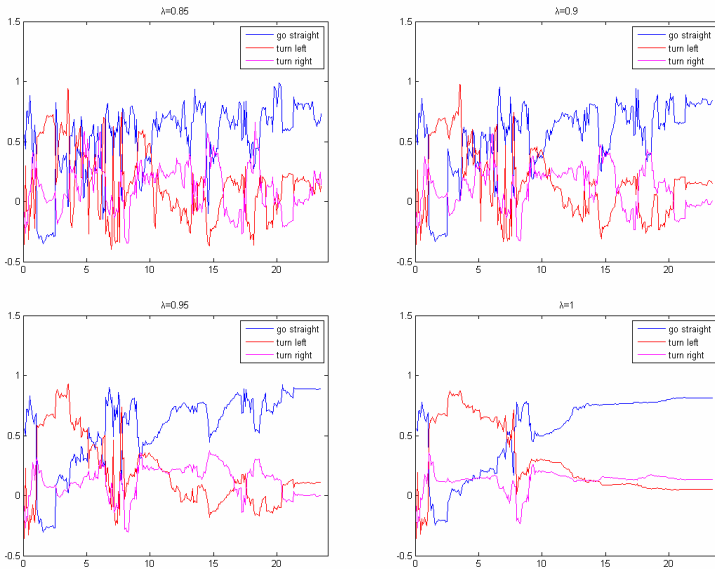


Fig. 3. Parameter estimations with different λ values. The smaller λ is, the more severe the estimation result fluctuates because when λ is small, the historical data puts less effect on the current estimation, that is, the estimation result is easy to fluctuate due to the newly coming data. When $\lambda=1$, the curve presents more stationary every historical data carries the same weight in the estimation and one newly coming data can hardly influence the estimation result.

Figure 4 gives different turn ratio curves of different links:

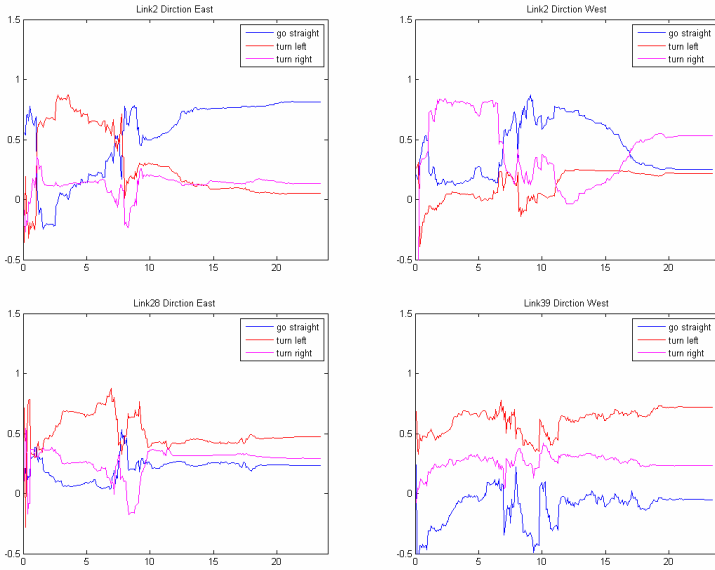


Fig. 4. Parameter estimations of different links. Every sub-figure shows significant parameter changes at about 8 am. These changes are corresponding to the rapid flow increase at 8 am. The rapid increase comes from the traffic habit of the citizens. Before 8, there is much less traffic than after 8 in Beijing. Before 8, the traffic flows are mostly bound for schools and household-related destinations while after 8, the traffic flows are mostly bound for office-related places. Thus makes the turn ratio of the same link vary between different time periods of the day. As for different links, the difference in turn ratio changes comes from the different surroundings of the links and the links' traffic capacity.

4.3 Cooperation with the RLS Algorithm with Correction

[14] provided a RLS turn ratio estimation algorithm with correction. With the same input data, this algorithm and ours give out different results:

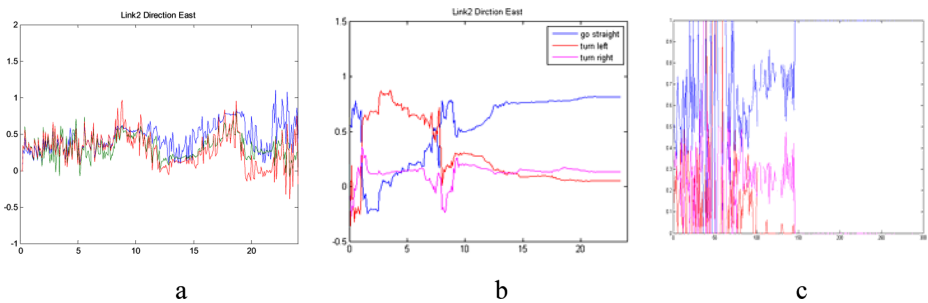


Fig. 5. a. RLS algorithm with correction b. RLS algorithm without correction c. actual curves

From Figure 5, we can see that our algorithm meets with the actual turn-ratios changing trend better. There is a trade-off between tracking speed and estimation reliability. Figure 5.c represents the highest tracking speed since the actual turn-ratios are derived only from the current flow data, taking no account of history data. Figure 5.a represents RLS algorithm with correction. Correction happens when the estimated turn-ratios changing obviously. Here the purpose of correction is to reduce the oscillation of estimation. But in RLS algorithm, the ‘forget determiner’ λ has already diminished the estimation oscillation. So this algorithm might cause over-correction.

4.4 Verification of Road Network Modeling

Build the road network model based on Adjacent Vector; and simulate the prediction of the entire network. Part of the result comes to Figure 6.

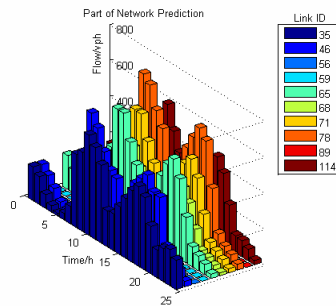


Fig. 6. Ten links’ prediction for 24 hours

Realize the entire prediction by Matlab, with computer configuration: Intel Core 2 Dou CPU, time using is listed in Figure 7.

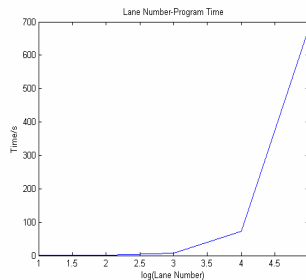


Fig. 7. Program time in different network scales

For normal networks, the magnitude of links is 10^3 , it will take less than 50 seconds to finish the prediction. Compared with the sample time-interval of 5 minutes in urban traffic system, this algorithm is advisable and capable to solve the real-time network prediction problems.

5 Conclusion

In this paper, we first construct the model of single-link based on Householder Transformation. In this way, we get a model with clearly meaningful parameters and relatively high accuracy. Then, we bring out a new network modeling structure of 'Adjacent Vector', which contains more information than the traditional adjacent matrix model and saves more memory space. Finally, we build a simulation traffic system based on real urban road backgrounds. In this model, our ideas prove to be efficient and advisable.

Acknowledgement. This work described in the paper is partially supported by National Natural Science Foundation of China (NSFC)50708054, Hi-Tech Research and Development Program of China (863Project) 2007AA11Z222, 2006AA11Z208, National Basic Research Program of China (973Project) 2006CB705506 and National Key Technology Research and Development Program 2006CBJ18B02.

References

1. Gao, H., Zhao, J., Jia, L.: Summary of Short-time Traffic Flow Forecasting Methods. *Journal of University of Jinan (Sci. & Tech.)* 22(1), 88–94 (2008)
2. Sun, X., Liu, T.: A Study on Urban Short-Term Traffic Flow Forecasting Based on a Nonlinear Time Series Model. *China Civil Engineering Journal* 41(1), 104–109 (2008)
3. Lee, S., Fambro, D.B.: Application of Subset Autoregressive Integrated Moving Average Model for Short-Term Freeway Traffic Volume Forecasting. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1678, Transportation Research Board of the National Academies, Washington, D.C., 179–188 (1999)
4. Wang, Z., Huang, Z.: An Analysis and Discussion on Short-Time Traffic Flow Forecasting. *Systems Engineering* 21(6), 97–100 (2003)
5. Wang, W., Dong, Y.: The Prediction Method of Short-Time Traffic Flow. *Journal of Shandong Jiaotong University* 12(2), 6–9 (2004)
6. Smith, B.L., Williams, B.M., Oswald, R.K.: Comparison of Parametric and Nonparametric Models for Traffic Flow Forecasting. *Transportation Research Part C: Emerging Technologies* 10(4), 303–321 (2002)
7. Wang, Y., Papageorgiou, M.: Real-Time Freeway Traffic State Estimation Based on Extended Kalman Filter: a General Approach. *Transportation Research Part B: Methodological* 39(2), 141–167 (2005)
8. Wang, L., Xu, H., Zheng, S.: The Study on Forecasting for Traffic Volume in ITS. *China Science and Technology Information*, pp. 268–269 (2008)
9. Hua, J., Faghri, A.: Dynamic Traffic Pattern Classification Using Artificial Neural Networks. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1399, Transportation Research Board of the National Academies, Washington, D.C., pp. 14–19 (1993)
10. Xu, N.: *Introduction to System Identification*. Electronic Industry Press, Beijing (1986)
11. Businger, P., Golub, G.H.: *Linear Least Squares Solutions by Householder Transformations*. *Numerische Mathematik* 7(3), 269–276 (1965)

12. Liu, Z.: On-Line Identification of a Time-variable System Based on Householder Transformation. *Signal Processing* 7(4), 220–227 (1991)
13. Zhang, H., Zhang, Y., Hu, D.: Study on Method of Traffic State Analysis for Urban Traffic Network. *Intelligent Transportation Systems* 6(1), 23–27 (2006)
14. Wang, Y.: Characteristics Extraction and Pattern Recognition of Typical Traffic Networks. Unpublished bachelor dissertation. Tsinghua University, Beijing (2008)

Research on Method of Double-Layers BP Neural Network in Prediction of Crossroads' Traffic Volume

Yuming Mao, Shiyang Shi, Hai Yang, and Yuanyuan Zhang

Department of Information Engineering, Shandong Jiaotong University, Jinan, China
maoyuming6096477@163.com

Abstract. Intelligent transportation systems(ITS) is effective on solving the problem of traffic jam in cities. Prediction of crossroads' traffic volume is the key technology in ITS. BP neural network is universally used in prediction of crossroads' traffic volume. This research aimed at using double-layers BP neural network to predict the traffic volume of Lishan Crossroad Jinan City. Results of the computer simulation showed that the method was applicable, the average relative tolerance was 9.71%. The double-layers BP neural network can be used for prediction of crossroads' traffic volume.

Keywords: Double BP neural network, crossroad, traffic volume, prediction.

1 Introduction

Along with the rapid development of economy and the increasing of vehicles in China, the crowding of city transportation, the rising of traffic accident rate, the low efficiency of the transportation, the serious situation of air pollution and the decreasing of the traveling safety, and so on, are more outstanding. Because of the limitation of the urban area, more and more cities begin developing intelligent transportation systems(ITS). ITS is effective on solving the problem of traffic jam in cities.

ITS is an accurate and efficient real time urban traffic management system with application of advanced information technology, data communication technology, electronic control and computer processing technology. Based on the researches home and abroad, the prediction of traffic volume is the key of ITS.

Prediction of traffic volume has a important station in ITS. There were some methods to predict the traffic volume at early period, the traditional prediction methods include autoregression model, autoregressive moving average model, moving average model and so on. These models are linear prediction models, them refresh data easily and compute fast. Because traffic volume has uncertainty and nonlinearity, these models' space of time is short, and the results are imprecise.

In recent years, some new models are used to predict traffic volume. The neural network is used to predict traffic volume has got some productions. The multi-layer forward artificial neural network is widely used recently, BP neural network is one of multi-layer forward artificial neural network, it is used to predict traffic volume, and the result is precise. In this research, the subsection learning of double-layers BP neural network was used to predict the traffic volume of Lishan Crossroad Jinan City.

2 BP Neural Network

BP neural network is error back propagation neural network, was published by Rumelhart in 1986. It is an adjusting artificial neural network with themselves toward potential relationship between input and output. BP neural network is a multilayer network, its neuron transfer functions are s type functions, they can accomplese any nonlinear mapping from input to output. The adjustment of weight uses back propagation algorithm. In the practical application of the artificial neural network, about 90% of the artificial neural network models adopt BP network or its change form, BP neural network applies to the approximation of function extensively, pattern recognition, the data compressed and so on.

BP neural network is three layers network, namely the input layer, the hidden layer and the output layer.

The neighboring layers fully connected by a weighting value, the units in the same layer are not connected with each other. A basic neuron has n inputs, each input connect with the next layer by weight W. Every layers weight can be adjusted by learning. The learning process is made up of model forword propagation and error back propagation. In forward process, the data inputed to the input layer propagate to the output layer through the hidden layer. The error propagate reversely and modify weights of each layer to minimize error. The following will be given the basic principle and derivation process.

Suppose the BP neural network consisting of three layers of node were trained by using the back propagation learning rule. The numbers of the input layer nodes, output layer nodes and hidden layer nodes were x_i , m_l and y_i respectively, the weight was w_{ij} between input layer nodes and hidden layer nodes, the weight was w_{ij} between output layer nodes and hidden layer nodes.

The following was a single neuron's output expression.

$$a = f (w \times p + b) \tag{1}$$

$$f (x) = \frac{1}{1 - e^{-x}} \tag{2}$$

f was input/output relation transfer function.

The following was output of hidden layer nodes.

$$y_j = f (\sum_i w_{ij} - \theta_j) \tag{3}$$

The following was output of hidden layer nodes.

$$m_l = f (\sum_j v_{jl} y_j - \theta_l) \tag{4}$$

Suppose that BP neural network had K layers, every layer has (m_1, m_2, \dots, m_k) neurons. There were H training sets (x_t, y_t) . The network training error of every training samples is the following.

$$E_K = \frac{1}{2} \sum_{k=1}^k \sum_{j=1}^{nl} (t_{jl} - y_{jl})^2 \tag{5}$$

$$E = \frac{1}{2N} \sum_{K=1}^N E_K \tag{6}$$

BP neural network amended the network weight on and on, until correction value was reduced to the acceptable range.

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} \tag{7}$$

$$w_{ij}(t + 1) = w_{ij}(t) + \Delta w_{ij} \tag{8}$$

$$\Delta v_{jl} = -\eta \frac{\partial E}{\partial v_{jl}} \tag{9}$$

$$v_{jl}(t + 1) = v_{jl}(t) + \Delta v_{jl} \tag{10}$$

η was learning rate at the above formulas.

Considering the conventional BP algorithm problems of slow convergence speed and easily getting into local dinky value, some improved BP algorithms are adopted in the practical application, one is the heuristic learning algorithm, the other is the training algorithm based on numerical optimization techniques, they have fast convergence speed.

3 Summarization on the Issue and Design of BP Neural Network

3.1 Summarization on the Issue

City roads are very complex, because there are many crossroads. It is significant to improve the traffic network's efficiency that increasing crossroad traffic volume and reducing the waiting time of every car. Nowadays, there are two ways of traffic signal timing plans: the first is settled program, and the entire period is fixed, the second is optimal timing plan. There are many optimal plans, autoconduction plan is a important optimal plan. In the plan, green light time contain maximum inductive green light time, the minimum inductive green light time and adding green light time. Commonly, under the roads laying coil detector near cross, when the green light time in the direction of AB over the minimum inductive green light time, and there is no vehicle passing, the direction of AB turns to no admittance condition, and the direction of CD turns to pass. If any vehicle to pass, it will increase a adding green light time stage until it reach the maximum inductive green light time, and then the direction of AB turns to no admittance condition, and the direction of CD turns to pass. In practise, we can use the two plans together. Using the first in rush hour, and using the second in other time.

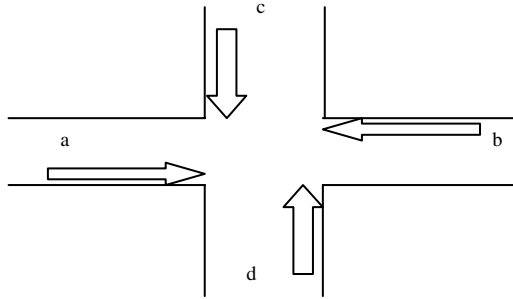


Fig. 1. Crossroad model

3.2 Design of BP Neural Network

The research designed a BP neural network to predict the crossroad’s traffic volume in future 5 minutes.

Firstly, single layer BP neural network model’s prediction precision is low, so the study adopted the double-layers BP neural network to predict traffic volume. Because the input and output layer neurons’ number was certain, the study needed determine the hidden layer neurons’ number. The study adopt trial and error method to determine the first and second hidden layer neurons’ numbers were 6 and 4 respectively.

Secondly, because the characteristic of BP neural network transfer function, equal proportion transform was adopted to limite the input and output data in the range from 0 to 1, that could improve precision.

Thirdly, the initial weights of BP neural network could use random number. Learning factor η decide weight change rate, a good learning factor is at the range from 0.1 to 0.6. In the study, η was 0.2. Momentum factor ∂ reduce concussion trend and improve convergence. In the research, ∂ was 0.9.

Considering the traffic volume of one direction of the crossroad related with the other directions’ traffic volume. In addition, the traffic volume of one direction of the crossroad related with the traffic volume of the adjacent crossroad in the same direction, so the research established time series prediction model. Suppose that $V_{ia}(t)$ was the traffic volume of the direction a of the crossroad i at t time section. The other directions’ traffic volume were $V_{ib}(t)$, $V_{ic}(t)$ and $V_{id}(t)$ respectively. Suppose that $V_{(i-1)a}(t)$ was the traffic volume of the direction i of the adjacent crossroad with crossroad a at t time section. $V_{ia}(t)$, $V_{ib}(t)$, $V_{ic}(t)$, $V_{id}(t)$ and $V_{(a-1)i}(t)$ were input samples, and $V_{ia}(t+1)$ was output sample.

4 The Application of BP Neural Network

The study used the traffic volume data of Lishan Crossroad Jinan City to predict traffic volume. The experiment data came from the historical data from June 25th, 2007 to July 1st, 2007. The study took the data one time each 5 minutes, there were 168 group data at Lishan Crossroad, used the first 148 group data to train the BP neural network and the latter 20 group data to test.

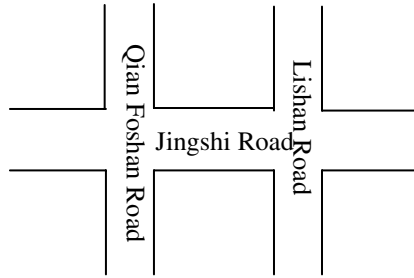


Fig. 2. Lishan Crossroad and Qian Foshan Crossroad

The experiment platform included the Langchao NL230DR server, Windows 2003 server operating system and Matlab 7.0 . Used Matlab 7.0 to train and simulate the designed BP neural network model. The traffic volumes of Lishan Crossroad from west to east direction in the t time segment, the $t-1$ time segment and the $t-2$ time segment, and Qian Foshan Crossroad from west to east direction in the t time segment were input samples. The traffic volume of Lishan Crossroad of road from west to east direction in the $t+1$ time segment is output sample. Prediction values and actual measurement values comparison and relative error as Fig. 3. and Fig. 4.

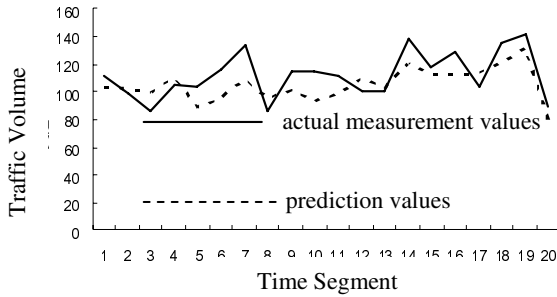


Fig. 3. Prediction values and actual measurement values comparison

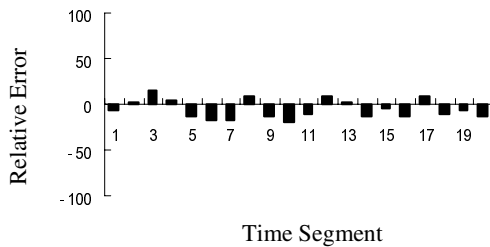


Fig. 4. Relative error

5 Conclusion

According to the result, the average relative error of the double-layers BP neural network model was 9.71%. The BP neural network model is feasible to predict traffic volume.

During predicting, the difficulty was how to determine the numbers of every layer's neurons and the learning factor and the momentum factor. In addition, the relative errors at the inflection point of the Fig.3. were big, that indicated the randomness of traffic volume was relative to other factors. In a word, the double-layers BP neural network model is feasible to predict traffic volume.

References

1. Liu, J.S., Fu, H., Liao, X.X.: Combination Prediction for Short-term Traffic Flow Based on Artificial Neural Network. In: WCICA 2006: Intelligent Control and Automation, pp. 8659–8663. IEEE Press, New York (2006)
2. Chen, S.Y., Wang, W., Ren, G.: A Hybrid Approach of Traffic Volume Forecasting Based on Wavelet Transform, Neural Network and Markov Model. In: IEEE International Conference on Systems, Man and Cybernetics, 2005, pp. 393–398. IEEE Press, New York (2005)
3. Shen, G.J., Sun, Y.X.: A Traffic Flow Model and Intelligent Control Technique for Urban Trunk Road. In: WCICA 2004: Intelligent Control and Automation, pp. 5321–5325. IEEE Press, New York (2004)
4. Morris, A.J., Zhang, J.: A Sequential Learning Approach for Single Hidden Layer Neural Networks. *Neural Networks* 11, 45–50 (2003)

Design and Implementation of the Structure Health Monitoring System for Bridge Based on Wireless Sensor Network

An Yin, Bingwen Wang, Zhuo Liu, and Xiaoya Hu

Department of Control Science and Engineering,
Huazhong University of Science and Technology, Wuhan 430074, China
ibm_hust@yahoo.com.cn, wangbw@public.wh.hb.cn,
liuzhuo@smail.hust.edu.cn, cindyhuxiaoya@sina.com

Abstract. For the shortcomings of the traditional wired way for structure health monitoring, a structure health monitoring (SHM) for bridge based on wireless sensor network (WSN) is proposed. The S-Mote node used for the WSN is designed and implemented which meets the specific hardware requirements of the structural health monitoring and supports the TinyOS as its operating system. The SHM system is deployed and tested on the ZhengDian Viaduct Bridge. In this deployment, 6 nodes are distributed over the main span, collecting the ambient vibrations at 100Hz. The collected data agrees with theoretical models and theoretical value of the bridge.

Keywords: WSN, SHM, S-mote, Deployment, ZhengDian viaduct bridge.

1 Introduction

Wireless sensor network (WSN) is composed of a large number of nodes which are equipped with sensor board and have the communicate function. The network is self-organizing. Multi-hop communication is used to transport the collected data to the base station. This allows random deployment in inaccessible terrains or disaster relief operations [1].

Recently the SHM based on wireless sensor network has become a hot research area. SHM system can estimate the state of structural health, detect damage or deterioration and determine the structure condition or health. SHM is a multidisciplinary research area that blending engineering knowledge in several areas, i.e., structural engineering, wireless technology, sensor technology and data analysis. The conventional method uses PC wired to piezoelectric accelerometers and wires have to run all over the structure. The drawbacks of such a system are high cost to installation, equipment and hard to maintenance. Compare to the conventional methods, the SHM based on WSN does not need any wiring. Thus the installation and maintenance are easy, inexpensive and less disruption of the operation of the structure [2, 3, 4].

In this paper, a structure health monitoring system for bridge is presented based on the self-developed node named S-Mote. The S-Mote node is equipped with

Micro-Controller (MSP430F1611), CC2420 and SD1221L acceleration sensor board. The paper will introduce the design part of the mote from hardware and software aspects. The later part of the paper will present the deployment on ZhengDian viaduct Bridge and analyze the collected acceleration data.

2 The Features of WSN in Structure Health Monitoring System

The wireless sensor network typically includes sensors nodes and base station. The base station not only collects data from sensor nodes, but also manages the network operation such as the distribution of command. The sensor nodes are the core of the WSN which collect data and transmit the data back to the base station. There are three main features of wireless sensor networks of structural health monitoring system.

(1) Energy

The capability of computing and communication of the sensor node is relatively weak and the energy equipped with the sensor node is limited. The widely used sensor nodes such as Mica2, Micaz, and IRIS are powered by two AA batteries. Since the power of sensor nodes are provided by the limited power that is not easy to change after the deployment, it is an important issue to extent the lift time of the sensor nodes in the wireless sensor network. In our self-designed S-Mote node, high-capacity power supply and low-power chip are used in hardware design framework.

(2) Nodes deployed manually

The deployment of sensor nodes in wireless sensor network can be classified into two different ways: randomly and manually. The latter is widely used in SHM in order to install the node in the key parts of the buildings. Once the nodes are deployed, they will not move during the monitoring process. As taking no account of the mobility of the nodes, the pre-configured routing and network topology can be used to transmit the packet.

(3) More complicated data analysis

Effective and accurate analysis of the data collected is the key part of the SHM system. It can evaluate the structure health state through analysis of the acceleration, strain and other key data. The collected data is valuable and not allowed loss.

3 The Design and Implementation of the Sensor Nodes

In this part, we introduce the hardware and software design according to the need of measuring the ambient vibrations of the bridge for structure health monitoring.

3.1 Hardware Design

For monitoring the structure health of Zhengdian viaduct Bridge, a new kind sensor node is designed named S-Mote. Compared to the existed sensor node such as mica, micaz, telosb, it is more suitable for structure health monitoring. The S-Mote is shown in Fig. 1. It consists of a mote (see Fig. 2), a sensor board (see Fig. 3) and two lithium batteries.

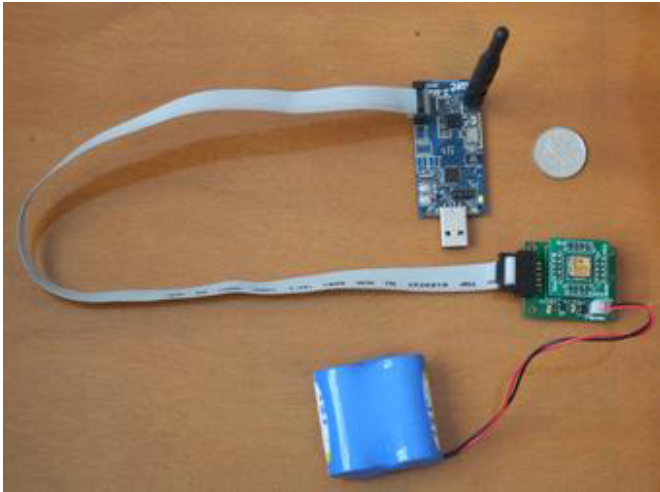


Fig. 1. The S-Mote node



Fig. 2. The mote designed for S-Mote

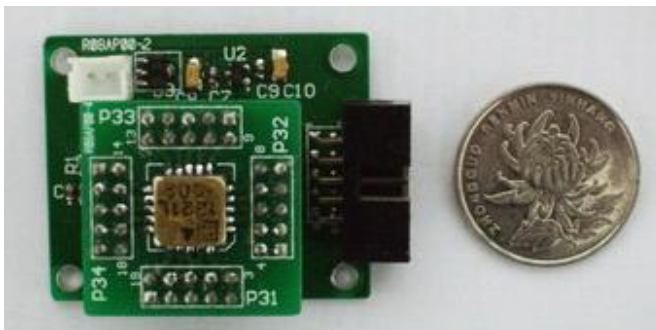


Fig. 3. The sensor board with SD1221L accelerometer

The S-Mote motes (see Fig. 2), which are used for SHM, are equipped with a Micro-Controller (MSP430F1611) which has 48KB of program memory, 10KB of RAM and provides processing, timer, watch dog component as well as analog-to-digital converters for sensor inputs, digital interfaces for connecting to other devices [5]. For storing the sample data, the M25P80 serial flash memory is used.

The S-Mote is also equipped with a RF tunable frequency radio chip(Chipcon CC2420) which runs at 2.4GHz providing 250Kbps bandwidth. The max work current in CC2420 is only 19.7mA makes it is more suitable for energy limited condition in WSN [6]. A new accelerometer board, shown in Fig. 3, is also designed used SD1221L accelerometer with a measurement range of $\pm 2g$ and system noise floor with $5\mu g / \sqrt{Hz}$ make it sensitive enough for bridge structure state monitoring [7].

The power supply of the lithium batteries whose power- capacity is 7500mAh are more powerful than the conditional AA battery. Because input voltage is 5V for SD1221L and the work voltage in mote only support 2.7-3.6V, a DC-DC converter is designed in the sensor board. The structure diagram of the S-Mote is shown in Fig. 4.

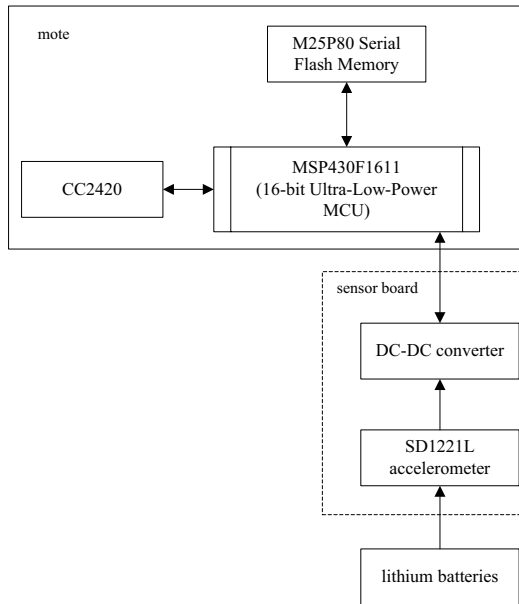


Fig. 4. The structure diagram of the S-Mote node

3.2 Software Designed

TinyOS is used in the S-Mote. TinyOS is an operating system developed by UC Berkeley and has been adopted by a large number of sensor network research group. It is a tiny (fewer than 400 bytes), flexible operating system built from a set of reusable components that are assembled into an application-specific system and support an event-driven architecture. NesC which is component- oriented language is used to program TinyOS code. It is similar with C language and support event-driven system [8, 9].

The program installed in the S-Mote is responsible for collecting the sample data during structure health monitoring, writing the sample data into the M25P80 serial flash memory and transmitting the data reserved in flash back to the base station. The work flow is shown in Fig. 5.

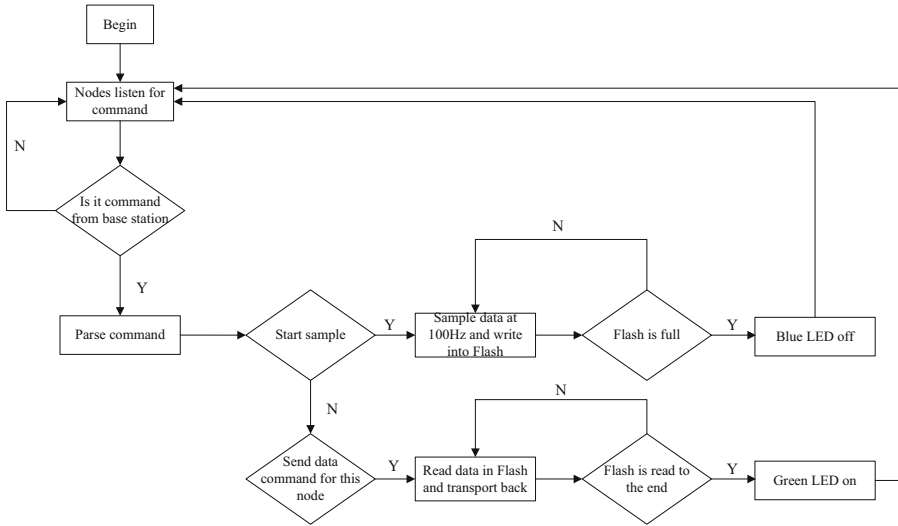


Fig. 5. The work flow of the S-Mote node

The green LED lights after the S-Mote nodes are deployed and are power-on which means the nodes are ready and waiting for the start command from base station. As soon as the nodes receive the start command, they collect the ambient vibrations data at 100Hz and write them back to the flash with the blue LED on. After the sample period, the nodes turn to sleep mode with blue LED off. The node reads the data reserved from flash and transports it back to the base station when it receives the send data command for itself.

3.3 The Monitoring Center System Software Design

The monitoring center system runs on laptop which connects the base station is responsible for controlling the network through sending different kinds of control command, storing the data in database and analyzing them. The architecture of the monitoring center system is shown in Fig. 6.

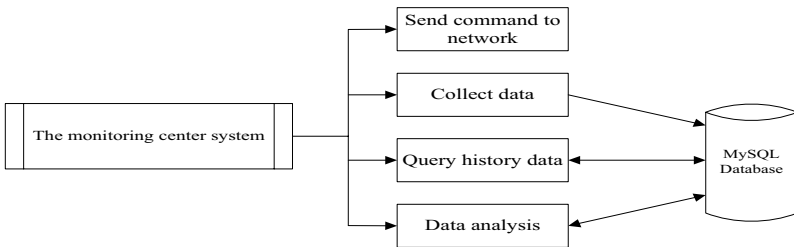


Fig. 6. The architecture of the monitoring center system

4 Deployment and Test

In order to test the S-Mote node and the performance of the software, a WSN for SHM is deployed on the ZhengDian viaduct Bridge. The ZhengDian viaduct Bridge is located in JiangXia near Wuhan city. The bridge is 1588ft (484m) long, 18 span. The WSN was designed with low cost and without interfering the operation of the bridge to measured ambient structural accelerations from traffic load. The goal was to evaluate the state of the bridge and compare actual behavior to design predictions. 6 nodes were deployed in the middle of the bridge with linear array (see Fig. 7).

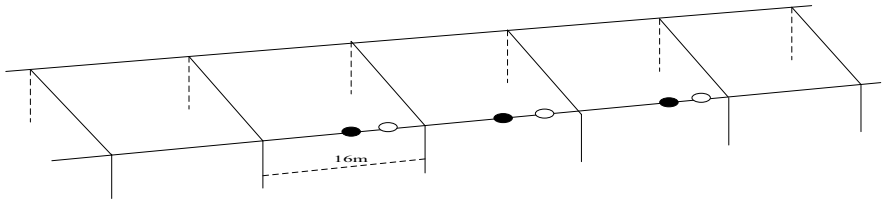


Fig. 7. The deployment of the WSN (the black nodes are deployed in the middle of the span and the white nodes are deployed in the 1/4 of the span)

The sensor board of the node was pasted on the surface of the bridge through rubber cement to maintain the SD1221L accelerometer in horizontal position. The base station was connected to a laptop which runs the monitoring center system program. The WSN for SHM is operating as follows. At the trigger signal from the base station, every node starts sampling the vibration data in 100Hz and fills up the data in the M25P80 serial flash memory on the S-Mote. Then the sampled data was transmitted back to the base station. This operation takes about 1.5 hours. The acceleration value in vertical orientation versus sample point of node 1 is shown in Fig. 8. It shows the amplitude with peaks of approximately 28mg, which corresponds to the passing of large cars.

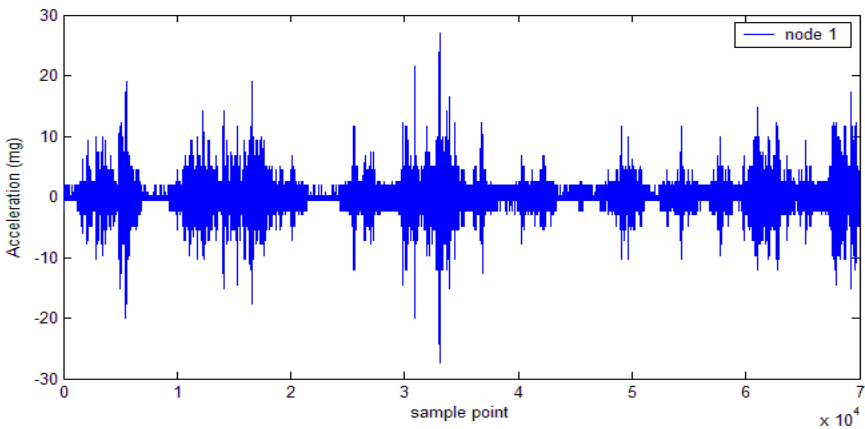


Fig. 8. The acceleration of the vertical orientation versus sample point

The fundamental tool of signal processing is the FFT. We use it to get the Power Spectral Density (PSD) to analyze the data. The algorithm is described as follows:

$$\hat{p}_{xx}(f) = \frac{|X_L(f_k)|^2}{f_s L} \tag{1}$$

$$f_k = \frac{kf_s}{N} \quad k = 0, 1, \dots, N - 1 \tag{2}$$

where $X_L(f_k) = \sum_{n=0}^{N-1} x_L[n]e^{-2\pi jkn/N}$, f_s is the signal sampling frequency, L is length of the signal. The Power Spectral Density of the sample data of node 1 is shown in Fig. 9. The frequency defined as H_1 correspond to the peak value of the power is 7.829Hz. The experimental value is 7.818Hz averaging H_i ($i \in [1, 6], i \in N$) of the six nodes deployed in the Zhengdian viaduct Bridge, which matches the theoretical value of 7.2Hz well.

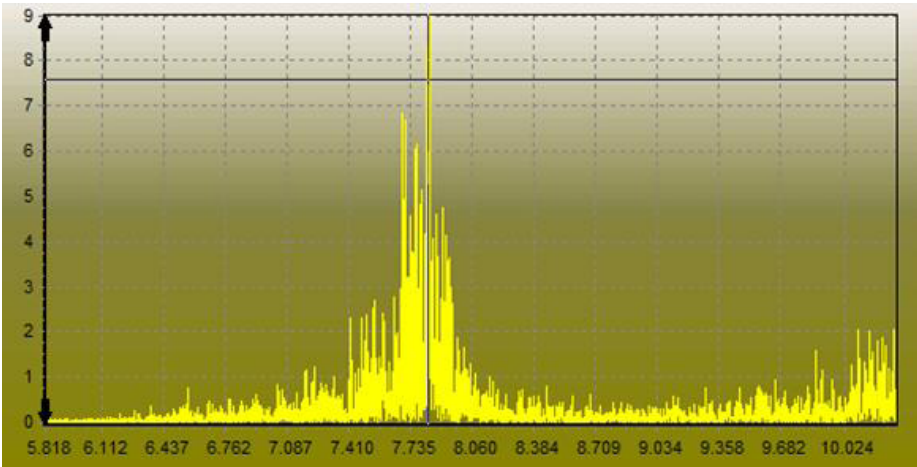


Fig. 9. The Power Spectral Density of the signals

5 Conclusion

In this paper, we present a structure health monitoring system for bridge based on wireless sensor network. A new node called S-Mote is designed and implemented which is equipped with SD1221L accelerometer to sample the subtle signals caused by ambient vibration. TinyOS, an operating system, is used in programming in the S-Mote node and the monitoring center system software is designed to control the network and analyze the sample data. 6 nodes were deployed in the middle of the ZhengDian viaduct Bridge and the experiment is done. The collected data agrees with theoretical models and theoretical value of the bridge. It turned out to be an effective way in SHM. Compared the conventional methods used the wired network in SHM, it is more easy to installation and less expensive.

Acknowledgments. The work was supported by the National Natural Science Foundation of China under the grant No. 60773190 and 60802002.

References

1. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless Sensor Networks: A Survey. *Computer Networks* 38, 393–422 (2002)
2. Hedley, M., Hoschek, N., Johnson, M.: Sensor Network for Structural Health Monitoring. In: 4th International Conference on Intelligent Sensors, Sensor Networks and Information Processing, pp. 361–366 (2004)
3. Kim, S., Pakzad, S., Culler, D., Demmel, J., Fenves, G., Glaser, S., Turon, M.: Health Monitoring of Civil Infrastructures Using Wireless Sensor Networks. In: 6th IEEE International Symposium on Information Processing in Sensor Networks, pp. 254–263 (2007)
4. Kinawi, H., Mahmoud, M., Naser, E.S.: Structural Health Monitoring Using the Semantic Wireless: A Novel Application for Wireless Networking. In: 27th Annual IEEE Conference on Local Computer Networks, pp. 770–780 (2002)
5. Chipcon Company, MSP430F1611 information, <http://focus.ti.com/docs/prod/folders/print/msp430f1611.html>
6. Chipcon Company, CC2420 Datasheet, <http://focus.ti.com/docs/prod/folders/print/cc2400.html>
7. SILICON DESIGNS, Inc., <http://www.silicondesigns.com/Pdf/1221.pdf>
8. Levis, P., Madden, S., Polastre, J.: TinyOS: An Operating System for Wireless Sensor Networks. *Ambient Intelligence*. Springer, New York (2004)
9. Gay, D., Levis, P., von Behren, P.: The nesC language: A Holistic Approach to Networked Embedded Systems. In: Proceedings of the ACM SIGPLAN 1998 Conference on Programming Language Design and Implementation, pp. 236–248 (1998)

Saving Energy in Wireless Sensor Networks Based on Echo State Networks

Ling Qin¹, Rongqiang Hu², and Qi Zhang³

^{1,2} Department of Information Engineering, Wuhan University of Technology
Wuhan, 430070, P.R. China

³ The National Key Laboratory of EMC, Wuhan, 430064, P.R. China
{qinling, hurongq} @whut.edu.cn

Abstract. Prolonged lifetime, robustness and scalability were important requirements in Wireless Sensor Networks (WSNs). We investigated the problem of energy-saving in Wireless Sensor Networks using Echo State Networks (ESN). In this research field, one key factor of the problems was how to save energy efficiently in battery-driven sensor nodes. We tried to present an approach addressing these difficulties based on ESN learning information of these sensors' history status when only the data was available. Echo State Networks utilized incremental updates driven by new sensor readings and massive short memory with history inputs, thus varying communication rates can help save energy. We evaluated this method against those traditional approaches to save energy, and observe that the quality of the overall operation was comparable to the approaches. Therefore, the ability of Echo State Networks to prolong lifetime during the sensor network operation made this approach more suitable and applicable.

Keywords: Wireless sensor Networks-WSNs, Echo state Networks-ESN, Recurrent neural Networks- RNN, Energy saving.

1 Introduction

With an increasing development in the research field of Wireless Sensor Networks (WSNs), more and more new kinds of applications were discussed in many platforms, like habitat or health monitoring, industrial, transportation systems automation and so on in the last decade. Prolonged lifetime, robustness and scalability played three leading and important roles in the applications. Especially, energy-saving in many real situations was crucial since battery-driven sensor nodes were severely energy-constrained. Considerable research has been recently carried out in an effort to make sensor network energy-efficient or energy-saving. In [2], a mathematical model was presented to determine a bound on sensor network lifetime, with and without sensing activities. The hardware-based energy model for transmission and reception described in [1] is widely used as the basic energy consumption model for a wireless sensor network node. Heinzelman et al. [7] proposed a cluster-based routing algorithm called LEACH as an energy-efficient communication protocol for wireless sensor

networks. The self-selected cluster heads collect raw data from the neighboring sensing nodes, aggregate them by data fusion methods, and transmit the aggregated data back to base stations for higher level processing. PEGASIS, an improvement over LEACH, was another example of an energy-aware protocol [11], which tends to increase the sensor network lifetime by decreasing the bandwidth via local collaboration among nodes. Another example was the TEEN protocol proposed in [6]. Dynamic power management [4] had also been used for the design of energy-efficient wireless sensor networks. Other related work includes energy-saving strategies for the link layer [12], data aggregation [9], and system partitioning [10].

In this paper we investigated the problem of energy-saving in Wireless Sensor Networks using Echo State Networks. Networks comprised various sensor nodes in WSNs had been regarded as a complex dynamic system, where future values depend on current and past values. Methods like feed-forward neural networks, employed for processing time series data however implicitly assumed a functional dependency between their current input and output. To consider a history of inputs, previous values had to be buffered outside the mechanism or explicitly encoded into it. Other approaches, such as recurrent neural networks, were able to compute outputs as a function over the input history, but had traditionally only rarely been used in practice due to long training times and intricate set up. Furthermore, many approaches required all input data of a time step to be present before it was possible to compute output values, which was in conflict with sensor network methodology where new data could arrive at any time.

To address these practical and methodological questions, we tried to present an approach for analyzing history inputs of sensor nodes in WSNs based on ESN [11]. ESN were a particular type of recurrent neural networks and had the ability to model dynamic systems. Unlike common recurrent neural networks, ESN did not suffer from the problem of slow convergence in training.

We showed that our approach using ESN with an explicit of the history of sensor inputs. The main contributions of this paper were as follows:

- We presented an approach to analyze how to save energy or prolong lifetime efficiently in Wireless Sensor Networks using Echo State Networks (Section 3).
- We evaluated our approach in the domain introduced above against compared three customary nonlinear equalization methods, namely a linear decision feedback equalizer (DFE), which is actually a nonlinear method; a Volterra DFE; and a bilinear DFE.
- We provide results of extensive experiments using ESN for sensors nodes under conditions with lower amounts of memory, computation and training data used. These results have practical implications for the use of ESN in the application of WSNs.

Section 4 concluded the paper.

2 Wireless Sensor Networks Topology

The topology of Wireless Sensor Networks generally consisted of a data acquisition networks and a data distribution networks, monitored and controlled by Basic Station Controller.

Gains in WSNs, sensor nodes transmitted and received various information, had been divided into different locations by they were located in areas. In each location one main gain called master gain, and the others called slave gain. In fact the master gain in one location could be main transceiver and digital signal processor, which have been in

charged of gathering local information, analyze data transmitted from other slave gains, transmit information to and receive commands from BSC (Basic Station Controller). With wireless communication as CDMA, GSM and Cellular network, users could use PDA, PC equipped with wireless transceiver, or Mobile phone to control or monitor the sensor networks.

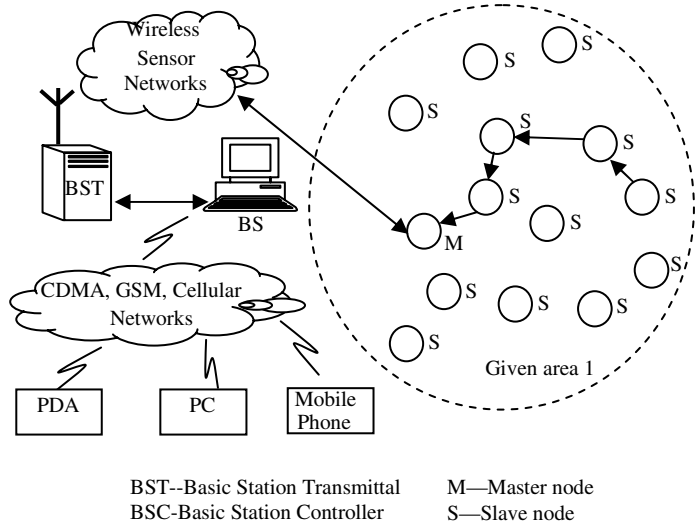


Fig. 1. Wireless Sensor Networks topology

3 History Inputs Analyzing Based on ESN

ESN have been developed from a mathematical and engineering perspective, but exhibit typical features of biological RNNs: a large number of neurons, recurrent pathways, sparse random connectivity and local modification of synaptic weights. The idea of using randomly connected RNNs to represent and memorize dynamic input in network states has frequently been explored in specific contexts.

The ESN approach differs from these methods in that a large RNN is used (order of 50 to 1000 neurons, previous techniques typically use 5 to 30 neurons) and in that only the synaptic connections from the RNN to the output readout neurons are modified by learning (previous techniques tune all synaptic connections). Because there are no cyclic dependencies between the trained readout connections, training an ESN becomes a simple linear regression task.

A possible method analyzed history inputs of sensor nodes was recurrent neural networks, as these are able to model dynamic non-linear systems. Moreover, WSNs had been often viewed as a typical dynamic non-linear system in applications. However, traditional recurrent neural networks have only rarely been used in practice, because of the slow convergence on training methods even for a small number of neurons.

3.1 Brief Formal Description on ESN

More formally, an ESN consisted of K input units, N internal units and L output units. Then, activations of input, internal, and output units at time step t were $u(t) = \{u_1(t), \dots, u_k(t)\}$, $x(t) = \{x_1(t), \dots, x_N(t)\}$, and $y(t) = \{y_1(t), \dots, y_L(t)\}$, respectively. Connection weights between units are kept in four connection matrices. There are $K \times N$ weights in the input weight matrix $W^{in} = (w_{ij}^{in})$, $N \times N$ weights in the internal weight matrix $W = (w_{ij}^{in})$, $L \times (K + N + L)$ weights in the output weight matrix, and $L \times N$ weights in a matrix $W^{back} = (w_{ij}^{back})$ for connection projecting back from the output to internal units.

The activation of internal units was calculated as

$$x(t+1) = f(W^{in}u(t+1) + Wx(t) + W^{back}y(t)) \tag{1}$$

with $f = \{f_1, \dots, f_N\}$ the output functions of the internal units – a sigmoid function for the experiments in this paper. Similarly, the output was computed as

$$y(t+1) = f^{out}(W^{out}(u(t+1), x(t+1), y(t))) \tag{2}$$

With $f^{out} = \{f^{out}_1, \dots, f^{out}_L\}$, the output functions of the output units and $\{u(t+1), x(t+1), y(t)\}$ the concatenation of input, internal, and previous output activation vector [7].

3.2 Motivation on ESN

First we created a random RNN with 1000 neurons (called the "reservoir") and one output neuron. Importantly, the output neuron was equipped with random connections that project back into the reservoir. A 3000 step teacher sequence $d(1), \dots, d(3000)$ was generated from the MGS equation and fed into the output neuron. This excited the internal neurons through the output feedback connections. After an initial transient, they started to exhibit systematic individual variations of the teacher sequence.

The fact that the internal neurons display systematic variants of the exciting external signal is constitutional for ESN: the internal neurons must work as "echo functions" for the driving signal. Not every randomly generated RNN has this property, but it can effectively be built into a reservoir.

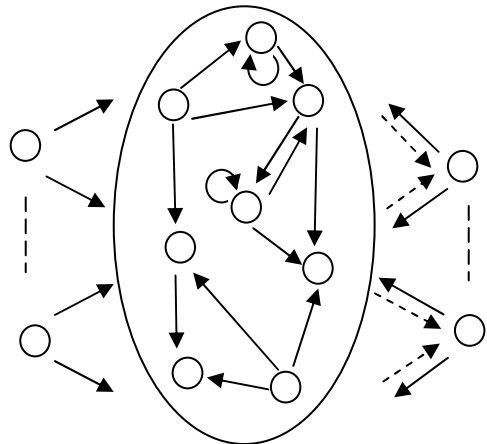


Fig. 2. Basic Echo State Networks architecture

It is important that the "echo" signals be richly varied. This was ensured by a sparse interconnectivity of 1% within the reservoir: this condition lets the reservoir decompose into many loosely coupled subsystems, establishing a richly structured reservoir of excitable dynamics.

3.3 Overview of Our Approach

The main reason for the jump in modeling accuracy is that ESN capitalize on a massive short-term memory. The basic idea for our approach was as follows: using the available data, we created a dynamical model of sensor nodes. We showed analytically [13] that under certain conditions an ESN of size N may be able to "remember" in the order of the last N inputs. This information was more massive than the information used in other techniques. We carried out numerous learning trials [14] to obtain ESN equalizers, using an online learning method (a version of the recursive least square algorithm known from linear adaptive filters) to train the output weights on 5000 step training sequences. We chose an online adaptation scheme [15] here because the methods in were online adaptive, too, and because wireless communication channels mostly are time-varying, such that an equalizer must adapt to changing system characteristics. The entire learning-testing procedure was repeated for signal-to-noise ratios ranging from 12 to 32 db. Using an ESN for nonlinear channel equalization: results. Plot showed signal error rate SER vs. signal-to-noise ratio SNR. **a.** linear DFE, **b.** Volterra DFE, **c.**

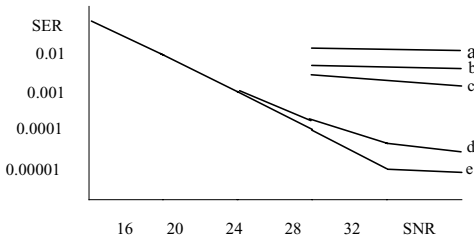


Fig. 3. SER and SNR based on ESN

bilinear DFE (a – c taken from [16]). d. represented average ESN performance with randomly generated "reservoirs" (error bars: variation across networks). **e.** indicated performance of best network chosen from the networks averaged in **d** (error bars: variation across learning trials).

During runtime, useless nodes were computed as deviation of the incoming data from predictions of the model. The initial threshold for deviation had to set by site personnel; once statistics of standby and working sensor nodes have been collected, a threshold for a specified standby rate can be set using the Neyman-Pearson test:

$$\lambda = \frac{\text{standby}}{\text{working}}$$

4 Conclusions and Future Work

We presented a systematic evaluation of energy-saving technique for WSNs based on ESN. Examples are varying sample and communication rates, detection of unusual behavior of environment or sensor network hardware, or compression of data. Varying communication rates can help save energy in WSNs.

Concluding, we believe ESN can help to solve energy-saving in WSNs. For future work, we plan to expand on the idea of using ESN in clusters of sensor nodes, using data from different domains.

References

1. Qin, L., Hu, R.: Intelligent Gain System Based on Complex and Dynamical Networks Model. In: 2007 IEEE International Conference on Robotics and Biomimetics, ROBIO, pp. 2018–2022 (2008)
2. Bhardwaj, M., Garnett, T., Chandrakasan, A.P.: Bounding the Lifetime of Sensor Networks Via Optimal Role Assignments. *IEEE Infocom* 3, 1380–1387 (2001)
3. Varshney, P.K.: *Distributed Detection and Data Fusion*. Springer, New York (1996)
4. Sinha, A., Chandrakasan, A.: Dynamic Power Management in Wireless Sensor Networks. *IEEE Design and Test of Computers* 18, 62–74 (2001)
5. Shah, R.C., Rabaey, J.: Energy Aware Routing for Low Energy Ad Hoc Sensor Network. In: *IEEE Wireless Communications and Networking Conference*, vol. 1, pp. 350–355 (2002)
6. Manjeshwar, A., Agrawal, D.P.: TEEN: A Routing Protocol for Enhanced Efficiency in Wireless Sensor Networks. In: *The 15th Parallel and Distributed Processing Symposium*, pp. 2009–2015 (2001)
7. Heinzelman, W.B., Chandrakasan, A., Balakrishnan, H.: Energy-Efficient Communication Protocol for Wireless Micro Sensor Networks. In: *The 33th Annual Hawaii International Conference on System Sciences*, pp. 3005–3014 (2000)
8. Heinzelman, W.B., Chandrakasan, A., Balakrishnan, H.: An Application-Specific Protocol Architecture for Wireless Microsensor Networks. *IEEE Transactions on Wireless Communications* 1, 660–670 (2002)
9. Wang, A., Heinzelman, W.B., Chandrakasan, A.P.: Energy-Scalable Protocols for Battery-Operated Micro Sensor Networks. In: *IEEE Workshop on Signal Processing Systems*, pp. 483–490 (1999)
10. Wang, A., Heinzelman, W.B., Chandrakasan, A.P.: An Energy-Efficient System Partitioning for Distributed Wireless Sensor Networks. In: *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, pp. 905–908 (2001)
11. Jaeger, H.: *The Echo State Approach to Analyzing and Training Recurrent Neural Networks*. GMD-Germany National Research Institute for Computer Science, GMD Report 148 (2001)
12. Shih, E., Calhoun, B.H., Seong, H.C., Chandrakasan, A.P.: An Energy-Efficient Link Layer for Wireless Micro Sensor Networks. In: *IEEE Computer Society Workshop on VLSI*, pp. 16–21 (2001)
13. Elfes, A.: Occupancy Grids: A Stochastic Spatial Representation for Active Robot Perception. In: *The 6th Conference on Uncertainty in AI*, pp. 60–70 (1990)
14. Jaeger, H.: *The Echo State Approach to Analyzing and Training Recurrent Neural Networks GMD-Report 148*. German National Research Institute for Computer Science (2001)
15. Mathews, V.J., Lee, J.: *Advanced Signal Processing: Algorithms, Architectures, and Implementations*. In: *Proc. SPIE, San Diego, CA*, vol. 2296, pp. 317–327 (1994)
16. Jaeger, H.: *Advances in Neural Information Processing Systems 15*. In: Becker, S., Thrun, S., Obermayer, K. (eds.), pp. 593–600. MIT Press, Cambridge (2003)

Enlargement of Measurement Range in a Fiber-Optic Ice Sensor by Artificial Neural Network

Wei Li, Jie Zhang, Ying Zheng, and Lin Ye

Department of Control Science & Engineering,
Huazhong University of Science and Technology,
Wuhan, 430074, China
vivily.li@gmail.com

Abstract. Artificial neural network (ANN) is employed to present a fiber-optic ice sensor (FOIS) with wide measurement range. Comparing with existing FOIS signal processing methods, this approach is not limited by the double-valued problem of output curve. Instead, it performs a measurement range from front-slope areas to back-slope areas. Moreover, this approach also handles the nonlinear problem of the sensor. As an application of the ANN, a calibration experiment platform is set up. The training samples are employed to train the ANN, and the testing samples are applied to surveil the predict ability of the ANN. The results obtained demonstrate the applicability of the proposed approach.

Keywords: Fiber-optic ice sensor, Measurement range, Neural networks, Double-valued.

1 Introduction

Fiber-optic ice sensor is a measuring instrument, which widely used in freezing environment. It offers a viable solution to the problem of detecting ice thickness and its rate directly. This kind of sensor has the advantages of simple structure, high precision, wide frequency response, immunity to electromagnetic interface, low cost, and small size. It can be used in a variety of applications with a little change, not only as an ice thickness sensor, but also as a secondary transducer for measuring physical properties such as temperature [1], pressure [2] and displacement [3]. Thus, fiber-optic ice detection is the key to a well developed fiber-optic detection technology, and some results have been obtained [4, 5, 6].

The performances of the FOIS, which depend on its material and geometric parameters, are presented in terms of output characteristic. The typical output characteristic, as shown in Fig.1, is the variation of sensor output voltage M with ice thickness d , $M=f(d)$.

It can be seen from this figure that the curve has front-slope, back-slope and peak. The voltage value increases in the front-slope areas but decreases gently in the back-slope areas. As two values of ice thickness (d_1 and d_2) correspond to one value of voltage M , so called double-value problem. We have to use only one slope areas to

detect the ice thickness and restrict the detection range. In addition, both slopes are nonlinear enough to produce influence on the precision of the ice sensor, all of which is similar to the characteristic of fiber-optic displacement sensor. Some authors have attempted to resolve those problems [7, 8]. However, their results were not satisfactory for the following aspects.

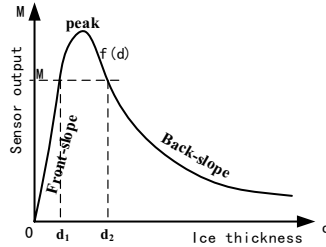


Fig. 1. The typical output characteristic of FOIS

- Accuracy: Some sensors use the switching components to enlarge the measuring range, which harm to the precision of the sensor.
- Expandability: The measurement range can not be enlarged enough from the front-slope areas to back-slope areas.
- Linearity: Although the measurement range can be enlarged, the nonlinear can not be coped with.

In this paper, a novel approach to enlarge the measurement range of the ice sensor is presented. The key point of the approach is the use of Artificial Neural Network (ANN), which processes the output signal to enlarge the measurement range from front-slope areas to back-slope areas, as well as calibrates the nonlinearity of the ice sensor in the range.

The paper is organized as follows. In section 2, the principle of FOIS with wide measurement range is presented. The design of probe is included. In section 3, based on a double-channel measurement, the problem of measuring range enlargement is addressed. An experiment study of the approach is provided in section 4. Concluding Remarks are collected in section 5.

2 Measurement Principle

The ice sensor has high resolution and sensitivity, however, its precision and sensitivity depend on the bundle front-end configuration. Typically, the arrangement of optical fibers in the bundle front-end may be random, concentric, hemi circular, double circular, concentric random, concentric hemi circular or hemi circular random. The schematic configuration of the ice sensor with wide measurement range is shown in Fig.2.

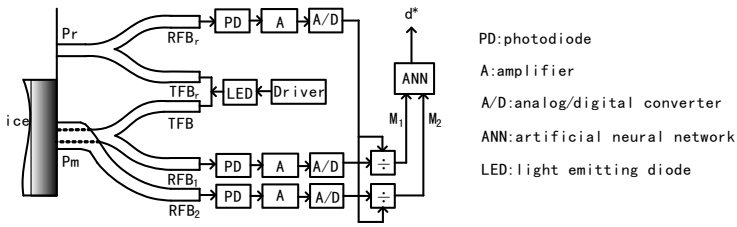


Fig. 2. The schematic configuration of the ice sensor with wide measurement range

The transmitting fiber bundle (TFB) and receiving fiber bundle 1 (RFB₁) are arranged in random order, while TFB and the receiving fiber bundle 2 (RFB₂) are arranged concentrically, as shown in Fig.3.

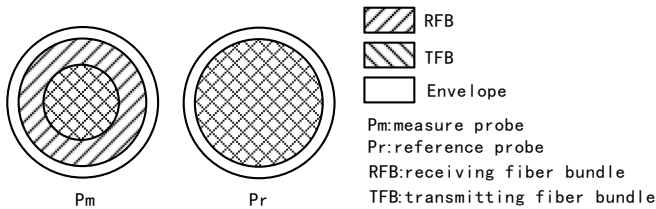


Fig. 3. The arrangement of probe

In order to eliminate the interference by background light, we employ a reference probe with random arrangement, which is electro-thermal to insure that there is no ice accretion on the probe. The light from LED is coupled into the reference transmitting fiber bundle (TRB_r) and TRB simultaneously. For analysis purpose, we have to take two cases into consideration.

- 1) light coupled into TFB_r
 In this case, there is no light modulated by ice available for the reference receiving fiber bundle (RFB_r). Only background light couples into RFB_r with intensity of I_r .
- 2) light coupled into TFB
 Light entering the ice volume is transmitted, scattered or reflected [9], and some of them go back along the RFB₁ and RFB₂, with intensity of I_{m1} and I_{m2} , respectively.

3 Neural Network Approach

As mentioned above, the photo-signals (I_r , I_{m1} and I_{m2}) in receiving fiber bundles are transformed into electric signals (V_r , V_{m1} and V_{m2}) by a photodiode, and then the electric signals are amplified and putted into a high speed AD converter. The digital signals obtained (V_r , V_{m1} and V_{m2}) can be expressed as follow:

$$V_r = D_r A_r C_r I_r \tag{1}$$

$$V_{m1} = D_{m1} A_{m1} C_{m1} I_{m1} \tag{2}$$

$$V_{m2} = D_{m2} A_{m2} C_{m2} I_{m2} \tag{3}$$

Where D_{m1} , D_{m2} , A_{m1} , A_{m2} , C_{m1} and C_{m2} represent the gain coefficient of photoelectric conversion, amplifier and AD conversion in the two measuring channel, respectively. D_r , A_r and C_r denote the gain coefficients in the reference channel.

We refer to the ratio of light coupled into TFB and light coupled into TFB_r as C_1/C_2 . I_r , I_{m1} and I_{m2} in Eq. (1)-(3) can be rewritten as:

$$I_{m1} = I_0 C_1 C_{01} \tag{4}$$

$$I_{m2} = I_0 C_1 C_{02} \tag{5}$$

$$I_r = I_0 C_2 \tag{6}$$

Where, C_{01} and C_{02} are conversion coefficients of the sensor. Substituting Eq. (4)-(6) into Eq. (1)-(3) yields:

$$V_{m1} = I_0 D_{m1} A_{m1} C_{m1} C_1 C_{01} \tag{7}$$

$$V_{m2} = I_0 D_{m2} A_{m2} C_{m2} C_1 C_{02} \tag{8}$$

$$V_r = I_0 D_r A_r C_r C_2 \tag{9}$$

Thus, according to Eq. (7)-(9), the output characteristic of sensor can be expressed as:

$$M_1 = \frac{V_{m1}}{V_r} = f_1(d)$$

$$M_2 = \frac{V_{m2}}{V_r} = f_2(d)$$

Therefore, compensation is implemented by the reference signal V_r . However, $f_1(d)$ and $f_2(d)$ are double-valued functions, with only one of which the detection range can not be enlarged from front-slope areas to back-slope areas. It should be noted that the locations of the peak in $f_1(d)$ and $f_2(d)$ are different due to the different arrangements employed by the measure probe as shown in Fig.3. There is an intersection point of $f_1(d)$ and $f_2(d)$. When use $f_1(d)$ together with $f_2(d)$ as output parameters, the input parameter is one-to-one correspondence with them, which can be expressed as:

$$d = g(M_1, M_2)$$

Where $g(*)$ is a nonlinear function that is difficult to be expressed by Mathematical Modeling.

In accordance with the high nonlinear mapping capability of the artificial neural network (ANN), various complex problems have been solved [10]. On the other hand, multilayered feed-forward networks have a better ability to learn the correspondence between input patterns and teaching values from many sample data by the error back propagation (BP) algorithm [11]. In this study, we employ a three-layered feed-forward neural network and trained it by error BP algorithm to build reverse model of the ice sensor. Fig.4 shows the architecture of the neural network that is considered.

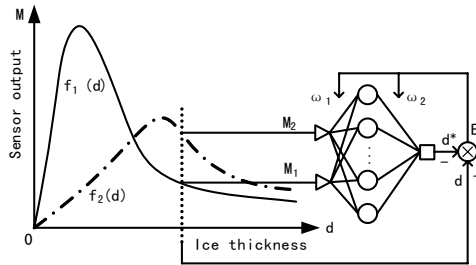


Fig. 4. The architecture of the neural network

The output voltage values of sensor (M_1 and M_2) are selected as input pattern. Ice thickness d^* is the output of the neural network. The error between the actual network output d^* and desired output d obtained in the calibration experiment are used to update the weight values ω_1 and ω_2 . The squared error function E_d is defined by:

$$E_d = \sum_{i=1}^m [d(i) - d^*(i)]^2$$

Where m represents the number of samples. The purpose is to make E_d small enough by choosing appropriate ω_1 and ω_2 , the effect of which is:

$$d^* \approx d = g(M_1, M_2)$$

After the training phase, the ANN parameters are fixed, and the output value d^* is enough close to the desired output d , so as to enlarge the measurement range from front-slope areas to back-slope areas and calibrate the nonlinear of the sensor output characteristics.

4 Experiment and Application

The algorithm proposed here has been trained and tested on an experiment platform. Both of the actual network output d^* and desired output d can be obtained. Fig.5 shows a simplified diagram of the experiment platform.

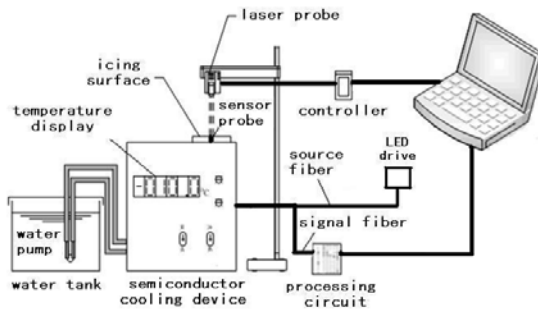


Fig. 5. Simplified diagram of the experiment platform

It can be seen from this figure that the platform consists of a semiconductor cooling device, a laser probe, a fiber-optic ice sensor, and a processing circuit. The semiconductor cooling device works just as an icebox to provide icing surface. The laser probe measures the desired ice thickness d with high precision. The fiber-optic ice sensor captures the optical information in ice volume. The processing circuit transforms the light intensity signal into voltage signal, and the division operation in it is completed by a PC computer. In order to build the neural network reverse model of the sensor, a calibration experiment is performed to obtain the inputs M_1 and M_2 , actual network output d^* and desired output d . A two-layer neural network with six hidden neurons and one output neuron is used. The error BP algorithm is employed to train the weights. The experiment consists of 1000-cycle training runs on the neural network, and the sun of squared error function is 1.93×10^{-6} . The neural network's software is written by the investigators.

Once the training finished, the trained neural network can be tested with other samples, and then it can be thought of as an "expert" in the ice thickness measurement. Our experimental results are shown in Fig.6, it is shown that the detection range is small with one measuring channel. Take curve M_2 as an example, if we use the front-slope for measuring, the detection range is limited to 0-2.7mm. On the other hand, the detection range is limited to 2.7-5mm in the back-slope. In addition, there is nonlinear relation between the ice thickness and sensor output.

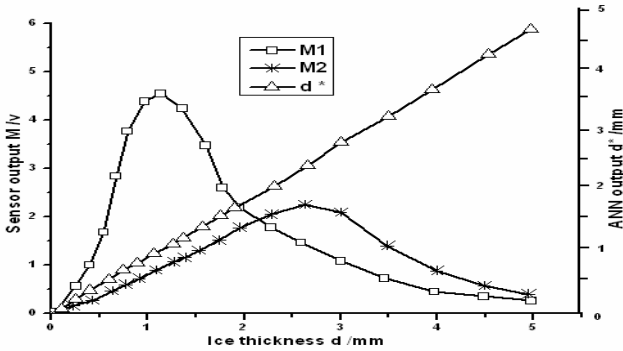


Fig. 6. Variation of sensor outputs M_1 , M_2 and neural network output d^*

Table 1. Some test samples of FOIS

	1	2	3	4	5
Sensor output M_1	0.03495	1.00601	2.84435	4.39241	4.24358
Sensor output M_2	0.00285	0.27089	0.59858	0.90005	1.16466
Actual displacement d	0.04331	0.39338	0.712	1.03574	1.36716
ANN output d^*	0.044	0.398	0.710	1.041	1.36
	6	7	8	9	10
Sensor output M_1	2.60841	1.787	1.09448	0.45221	0.27089
Sensor output M_2	1.52185	2.05598	2.09721	0.8894	0.40197
Actual displacement d	1.75299	2.28255	3.00289	4.002	4.96223
ANN output d^*	1.751	2.291	3.002	4.004	4.978

Some experiment measure data are shown in Table 1. The results show that a smooth transition from front-slope areas to back-slope areas is realizable for the neural network output. In this study, with both the measuring channels, we demonstrate the increase in the measurement range from front-slope to back-slope by using ANN to process the output signal of the ice sensor.

5 Conclusion

In present paper, an ANN approach, which enlarges the measurement range of FOIS, is presented. Comparing with existing FOIS signal processing methods, our approach has the following advantages. First, it is not limited by the double-valued problem of output curve. Second, it performs a wide measurement range. Third, it handles the nonlinear problem. With a FOIS, whose fiber bundle arrangement are concentric random and concentric, the calibration experiment platform has been carried out.

Based on the calibration experiment data, the training samples are employed to train the ANN, and the testing samples are applied to surveil the predict ability of the ANN. The authors have satisfactorily come to the conclusion that this approach is feasible and makes a breakthrough in the traditional signal processing. However, ANN trained by BP algorithm has low convergence rate and high requirements on the sufficiency and validity of the samples. The future work is required to pay more attention to the two aspects: one is the purpose for optimizing ANN algorithm, the other is the detailed projects of improving experiment platform. Those are the directions of the future research work.

Acknowledgments. The authors are grateful for the financial support provided by the National Nature Science Foundation of China (No.10577008 and 60604030).

References

1. Meng, Q.: Fiber-optic Temperature Sensor Based on Semiconductor Optical Absorption. *J. Information Technology and Informatio* 5, 48–50 (2004)
2. Gong, H.: Reflective Optical Fiber Micro-displacement Sensor. *J. Journal of Transducer Technology* 22, 16–17 (2003)
3. Zhao, Z., Gao, Y., Luo, Y.: Fiber-optic Pressure Sensor. *J. Journal of Transducer Technology* 4, 13–15 (2005)
4. Ikiades, A., Glen, H., Armstrong, D.J.: Measurement of Optical Diffusion Properties of Ice for Direct Detection Ice Accretion Sensors. *J. Sensors and Actuators A: physical* 140, 24–31 (2007)
5. Ikiades, A., Armstrong, D.J., Howard, G.: Optical Diffusion and Polarization Characteristics of Ice accreting in Dynamic Conditions Using a Backscattering Fiber Optic Technique. *J. Sensors and Actuators A: Physical* 140, 43–50 (2007)
6. Kiades, A.: Direct Ice Detection based on Fiber Optic Sensor Architecture. *J. Applied Physics Letters* 91, 104–106 (2007)
7. Ma, J., Zhang, L., Tang, W.: The Reference Scheme of Optic Fiber Displacement Sensor: Analysis and Design. *J. Laser Journal* 17, 197–200 (1996) (in Chinese)

8. Zhang, H., Li, X.: Design of Electronical Circuit for the Linearity Compensation of a Reflective Type Fiber-optic Displacement Sensor. *J. Journal of Transducer Technology* 18, 19–22 (1999) (in Chinese)
9. Ikiades, A., Armstrong, D.J., et al.: Fiber Optic Sensor Technology for Air Conformal Ice Detection. *J. Industrial and Highway Sensors Technology* 5272, 357–368 (2004)
10. Sejnowski, T.J., Rosenberg, C.: Parallel Networks That Learn to Pronounce English Text. *J. Complex System* 1, 145–168 (1987)
11. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning Internal Representations by Error Propagation. In: *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, pp. 318–362. MIT Press, Cambridge (1986)

Epidemic Spreading with Variant Infection Rates on Scale-Free Network

Liu Hong^{1,2}, Min Ouyang^{1,*}, Zijun Mao¹, and Xueguang Chen¹

¹ Institute of System Engineering, Huazhong University of Science and Technology, Wuhan, 430074, P.R. China

² Key Lab. for Image Processing & Intelligent control, Huazhong University of Science and Technology, Wuhan, 430074, P.R. China

Abstract. In this paper, we proposed a susceptible-infected model with variant infection rates because different individuals have different resistance to diseases in different periods of real epidemic events. We consider two cases: Case 1, we know every individual's infection rate to a kind of epidemic, satisfy a type of distribution. Case 2, assume all individuals have same initial infection rates, a susceptible individual's infection rate will be less than the initial rate if he is not infected after limited number of contacts with infected ones. For both two cases, at the time t_D , preventive and control measures bring into effects, every individual's infection rate would decrease. We implemented this models on scale-free networks, and found that the epidemic process before the time t_D in Case 1 is almost the same as that in Standard SI model if the infection rate in Standard SI model equals the mean infection rate in Case 1. Furthermore, using numerical simulation, we analysis the effects of the parameter t_D , and find the bimodal distribution of final infection rates. Finally, we conclude what we get in this paper and give our future direction.

Keywords: Epidemic spreading, Variant infection rate, Scale-free network.

1 Introduction

The previous works about epidemic spreading in scale-free networks present us with completely new epidemic propagation scenarios in which a highly heterogeneous structure will lead to the absence of any epidemic threshold [1, 2]. These works mainly concentrate on the susceptible-infected-susceptible SIS [3, 4], susceptible-infected-removed SIR [5, 6]. Another typical model, susceptible-infected SI [7-8] models, also become the focus of the study, because in many cases, this model is more suitable to describe the dynamical process, for example, In the process of broadcasting [8], each node can be in two discrete states, either received or unreceived. A node in the received state has received information and can forward it to others like the infected individual in the epidemic process, while a node in the unreceived state is similar to the susceptible one. Since the node in the received state generally will not

* Corresponding author.

lose information. What's more, this model is more appropriate than SIS and SIR models when investigating the dynamical behaviors in the very early stage of epidemic outbreaks when the effects of recovery and death can be ignored.

However, the common assumption in most of the aforementioned works [9–11], when a susceptible individual contacts with an infected individual, the susceptible individual will be infected at a constant infection rate λ in the whole epidemic process. But in real system, individuals may have different infection rates and the infection rates may even change in time due to efficient preventive and control measures. It has been pointed out that the proper formulation of the infection rate requires taking many more detailed factors into account [12]. Under this consideration, they display that different infection rates lead to different threshold behaviors in the SF networks [12]. And then Ke Hu took into account the effect of density of infected neighbors around an individual in the definition of spreading rate, some interesting results were found [7].

In our paper, we thought the infection rate of a susceptible individual was determined by his own body mass and the preventive and control measures, this means special body mass might have a better resistance to a kind of epidemic, and efficient preventive and control measures would lead to lower infection rate for susceptible individuals (In our paper, the infection rate of a susceptible individual means the rate that this susceptible individual would be infected, maybe we use "infected rate" better, but there is no such as kind of usage). At the beginning of the epidemic, people know little about the epidemic, nothing preventive and control measures were taken, the infection rates should be different for different body mass. After a period of time t_D , people recognize the severity of epidemic and then more preventive and control measures will bring into effect, the infection rates should also be different. For example, the SARS happened in Hong Kong in 2003 [13], all Indian resided there are not infected, while children are more difficult to be tainted than adults, so are some special people. When people realized the severity of this epidemic, some corresponding measures were taken, the number of infected individuals in every day decreases rapidly. Another example is the spreading of Hepatitis B, when people know more about this epidemic. Although it can't be cured, some preventive and control measures are found, so many susceptible individuals will hardly be infected, we can say their infection rates tend to be zeros. Obviously, the infection rates are different for different people in different periods.

So a new SI models consider different infection rates for different people in different periods are more suitable to describe the epidemic process. In our paper, we will put forward two kinds of improvement models in terms of different considering on the variant infection rates.

2 Model

In the Standard SI network model [9–11], individuals can be in two discrete states, either susceptible or infected. A susceptible individual will be infected at rate λ when he contacts with an infected individual. The infected individual will be still infected in the whole process. The total population N is assumed to be constant. Thus, if $S(t)$ and $I(t)$ are the number of susceptible and infected individuals at time t , respectively, then $N = S(t) + I(t)$. In our model, we will consider individual's different infection rates, so we consider the two cases as follow:

Case 1: We know every individual's infection rates, satisfy a kind of distribution. The different infection rates are attributed to different body mass. At the time t_D , some preventive and control measures bring into effect, and then the infection rates will decrease. Assume the infection rate at the time t is λ_{before} , and then the infection rate at the time $t+1$ will be $\lambda_{after}=f_1(t, \lambda_{before})$.

Case 2: We don't affirm individual's infection rates in advance, but if a individual is still susceptible after m_1 contacts with infected ones, we can think he has relatively better resistance, corresponding to lower infection rate. Assume the infection rate of a susceptible individual is λ_{before} at the time t , after m_1 contacts with infected individuals, if he is still susceptible, then the infection rate will be $\lambda_{after}=f_2(m_1, t, \lambda_{before})$. At the time t_D , people know the severity of the epidemic and some preventive and control measures bring into effect, then the infection rate will be influenced by both body mass and preventive and control measures. Assume the infection rate at the time t is λ_{before} , after m_2 contacts with infected individuals, if he is still susceptible, then the infection rate will be $\lambda_{after}=f_3(m_2, t, \lambda_{before})$. In this case, the infection rates change in dependence of the time.

From the papers [14][15], we know that human body mass satisfies the normal distribution, and then we can use the normal distribution to reflect individual's different infection rates of different body mass. So we can adopt normal distribution in Case 1. What's more, before the time t_D in Case 2, assume the normcdf denotes the cumulative normal distribution function(cdf), $p = \text{normcdf}(\lambda, u, \delta)$ returns the cdf of the normal distribution with mean u and standard deviation δ ; while norminv denotes the inverse of the normal cumulative distribution function, $\lambda = \text{norminv}(p, u, \delta)$ returns the inverse cdf for the normal distribution with mean u and standard deviation δ , evaluated at the values in p . so the function f_1 can be expressed as $\lambda_{after}=f_2(t, m_1, \lambda_{before}) = \text{norminv}(\text{normcdf}(\lambda_{before}, u, \delta) - k_1 * m_1 / (1/\lambda_{before}), u, \delta) = \text{norminv}(\text{normcdf}(\lambda_{before}, u, \delta) - k_1 * m_1 * \lambda_{before}, u, \delta)$, here, $k_1 * m_1 * \lambda_{before}$ describe the probability that his infection rate is less than λ_{before} , $k_1 < 1$ is a gain parameter. In our experiment, set $k_1=0.1, m_1=10, u=\delta=0.01$;

After the time t_D , some measures will bring into effects, both body mass and preventive and control measures will affect the infection rates. So the function f_3 should be different from f_2 . From the SARS happened in Hongkong in 2003, when people and government took some measures to prevent the epidemic, the infection rate decreased, as can be seen in Fig. 1. The curves can be well approximated using the function $\lambda = a * \exp(-b * t)$, where $b(b < 1)$ is a gain parameter. In our paper, we will use this type of exponential function to reflect the decrease of individual's infection rates due to preventive and control measures, so the function f_3 can be described as $\lambda_{after}=f_3(m_2, t, \lambda_{before}) = \exp(-m_2 * k_2) * \lambda_{before}$, $k_2 < 1$ is a gain parameter. In our simulation, set $m_2=6; k_2=0.2/6$; we can also use this type of function in Case 1, so the function f_1 in our experiment will be described as $\lambda_{after}=f_1(t, \lambda_{before}) = 0.05 * \lambda_{before}$.

Next, we will discuss our above models from the theory, and then the simulation results will testify what we get and give us some new understanding about the dynamic process.

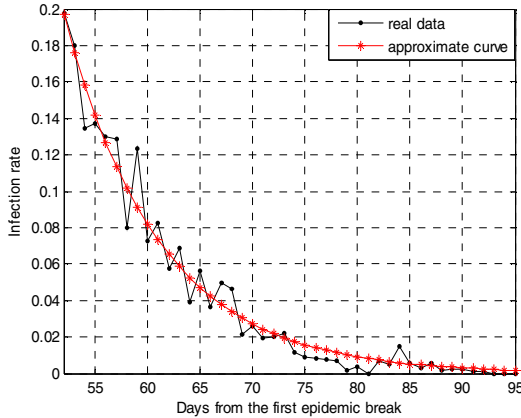


Fig. 1. The infection rate decreasing in dependence of the days about the SARS happened in Hongkong in 2003, The star-marker lines represent the approximation by the function $\lambda=a*\exp(-b*t)$, the best fit parameter are $a=60, b=0.11$

3 Theory Analysis

In this section, we try to analysis the differences between standard SI model with constant infection rates and Case 1 with variant infection rates for different body mass before the time t_D .

The network is represented by a connected graph $G(V,E)$, where V is the set of n individuals and E the set of relations between them. The state of the system, at time t , is described by a vector $X(t)=[x_1(t),x_2(t),\dots,x_n(t)]$, where $x_i(t)=1$ if the individual i is infected and $x_i(t)= 0$, otherwise. Denote by $A=(a_{ij})_{i,j=1,\dots,n}$ the adjacent matrix of the graph structure on the set of individuals. $a_{ij}=1$ if there is a link between node i and j , and $a_{ij}=0$, otherwise. Denote $y_i(t)$ as at time t the number of the i th individual's infected neighbors, then at time t , the i th individual can be infected at the rate $P(y_i(t))<1$ (k is a limited number), if we generate a uniform distributed random data ξ between 0 and 1, if $\xi<P(y_i(t))$, we can say this individual will be infected, if $\xi>P(y_i(t))$, he will still be susceptible. If we use the function $Ceil(x)$ to return the minimum integer not less than x , and then the i th individual's state can be expressed as $Ceil(P(y_i(t))-\xi)$. Because $0<P(y_i(t))<1, 0=<\xi<=1$, and then $-1<P(y_i(t))-\xi<1$, the value of $Ceil(P(y_i(t))-\xi)$ is 0 or 1. So the epidemic process of the Standard SI model can be described as follow.

$$[y_1(t), y_2(t), \dots, y_n(t)] = [x_1(t), x_2(t), \dots, x_n(t)]A \tag{1}$$

$$[z_1(t+1), \dots, z_n(t+1)] = Ceil([P(y_1(t)), \dots, P(y_n(t))] - rand(1, n)) \tag{2}$$

$$X(t+1) = Z(t+1) \vee X(t) \tag{3}$$

In the equation (2), the $rand(1, n)$ denote a row vector, it has n independent random element all between 0 and 1, $(z_i(t))_{i=1,\dots,n}$ denotes that at time t , the i th individual's state, but it contains the followed case: the i th individual who is infected at time $t-1$

may become susceptible at time t because $P(y_i(t)) < \xi$. The operation \vee denotes the “or” operation, if one of the two data is 1, and then the result is 1, if both two data are 0, and then the result is 0, so the equation (3) can ensure the infected individuals at time $t-1$ must be infected at time t . We can see the equations (1), (2) and (3) can describe the SI model equally.

For Case 1, equation (2) should be changed into equation (4) as follow:

$$[z_1(t+1), \dots, z_n(t+1)] = \text{Ceil}([P_1(y_1(t)), \dots, P_n(y_n(t))] - \text{rand}(1, n)) \tag{4}$$

Here $P_i(y_i(t)) = 1 - (1 - \lambda_i)^{y_i(t)}$, $\lambda_i (i=1, 2, \dots, n)$ ($0 < \lambda_i < 1$) satisfies a kind of distribution. Equation (4) is a random equation, so we can consider the expectation $E(Z(t))$, and then the equation (4) can be changed into equation (5).

$$E([z_1(t+1), \dots, z_n(t+1)]) = E(\text{Ceil}([P_1(y_1(t)), \dots, P_n(y_n(t))] - \text{rand}(1, n))) \tag{5}$$

First, we give a theorem.

Theorem 1. If A and B are independent random variables, B satisfies the uniform $[0, 1]$ distribution, and the variable $A \in [0, 1]$, then $E(\text{Ceil}(A-B)) = E(A)$.

Proof. (1) random variable A is continuous random variable.

Assume that the density function of random variable A is $f(a), a \in [0, 1]$, the density function of B is $f(b)=1, b \in [0, 1]$, so the joint density function of A and B is $f(a, b)=f(a)*f(b), a \in [0, 1], b \in [0, 1]$, then,

$$\text{Ceil}(A - B) = \begin{cases} 0 & A \leq B \\ 1 & A > B \end{cases}$$

$$E(\text{Ceil}(A-B)) = 1 * P(A > B) = \int_0^1 \int_0^a f(a) f(b) da db = \int_0^1 a f(a) da = E(A)$$

(2) random variable A is discrete random variable .

Assume that the value of random variable A is $a_1, a_2, \dots, a_n, 0 \leq a_1 < a_2 < \dots < a_n < 1$ and the corresponding probability is $p_1, p_2, \dots, p_n, 0 < p_i \leq 1, \forall i = 1, 2, \dots, n$

$$\begin{aligned} E(\text{Ceil}(A - B)) &= P(A > B) \\ &= \sum_{i=1}^n p_i P(B < a_i) = \sum_{i=1}^n p_i a_i = E(A) \end{aligned}$$

So the theory is right no matter the type of the random variable A . For small λ , $P_i(y_i(t)) = 1 - (1 - \lambda_i)^{y_i(t)} \approx \lambda_i y_i(t)$, and then, we can infer:

$$\begin{aligned} &E(\text{Ceil}([P_1(y_1(t)), \dots, P_n(y_n(t))] - \text{rand}(1, n))) \\ &= E([P_1(y_1(t)), \dots, P_n(y_n(t))]) \approx E([\lambda_1 y_1(t), \dots, \lambda_n y_n(t)]) \\ &E(\text{Ceil}([P(y_1(t)), \dots, P(y_n(t))] - \text{rand}(1, n))) \\ &= E([P(y_1(t)), \dots, P(y_n(t))]) \approx E([\lambda y_1(t), \dots, \lambda y_n(t)]) \end{aligned}$$

We know that λ_i and $Y_i(t)$ are independent random variables, so we can infer that:

$$E([\lambda_1 y_1(t), \dots, \lambda_n y_n(t)]) = \lambda E([y_1(t), \dots, y_n(t)])$$

So for small λ , if the mean of random variable λ in Case 1 equals the λ in standard SI model, the equation (2) and (4) can be thought equally. So the epidemic processes before the time t_D in Case 1 should be almost the same with that in Standard SI model.

4 Simulation Analysis

To demonstrate the differences of our improvement models on the spreading processes, we perform extensive numerical simulations on the Barabási and Albert (BA) networks[5], the network size $N=1000$, and in our paper, we consider three kinds of cases: Standard SI model, Case 1 and Case 2. In standard SI model, assume the constant infection rate $\lambda=0.01$; In Case 1, we adopt the normal distribution with mean $\mu=0.01$ and standard deviation $\delta=0.01$. Next, we will discuss our models from the several aspects as follow.

(1) Epidemic spreading processes

Barthélemy et al. [9,10] studied the SI model in Barabási-Albert (BA) scale-free networks, and found that the density of infected nodes, denoted by $i(t)$, grows approximately in the exponential form, $i(t) = e^{ct}$, where the time scale c is proportional to the ratio between the second and the first moments of the degree distribution, $c \sim \langle k^2 \rangle / \langle k \rangle$.

From the theory analysis in last section, the epidemic processes of Standard SI model and Case 1 before the t_D should be almost the same when the infection rate in Standard SI model equals the mean infection rate in Case 1. The similar result can be seen in fig.2 by means of simulation. We set $t_D=100$, and run our simulation for about 100 times, the average of all values was the final result. From the graph, the almost

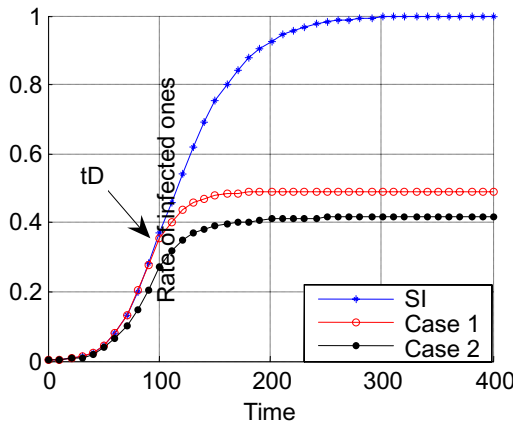


Fig. 2. Average rate of infected individuals in dependence of the simulation time for SI model, Case 1, Case 2. The infection rate in Standard SI model, mean infection rate in Case 1 and initial infection rate in Case 2 are all 0.01, $t_D=100$, other parameters settings are introduced in section 2. We run our simulation for about 100 times, the average value as the result.

same epidemic processes were shown before the time t_D , but after the preventive and control measures bring into effects, the process in Case 1 become slow, and reach the steady state quickly, some people would not be infected finally, this can explain why few epidemics can cause all individuals being infected.

However, in Case 2, the initial infection rates are all assumed to be 0.01, before the time t_D , some individuals were thought to have better resistance if they were still susceptible after a certain number of contacts with infected ones. These individuals have lower infection rates, and then mean infection rate decreased gradually before the time t_D . So the epidemic process for Case 2 is slower than that for Case 1, as can be seen in fig.2.

From the epidemic process, we know that: On the one hand, if we use the constant infection rate to study the statistical characters of epidemic processes instead of individual's random infection rate satisfying a kind of distribution whose mean value equal the constant infection rate in Standard SI model, we can also get the similar results. On the other hand, preventing and control measures should be taken as soon as possible, so the parameter t_D is important in our present model.

(2) Effect of the parameter t_D

As can be seen in fig.2, the parameter t_D has a large impact on the epidemic process, so in this section, we will discuss its effects from two aspects. First, we analysis the effects on the epidemic process. The number of increased infected ones in dependence of the time can be seen in fig.3, we fix the parameter $t_D=100$, and run our simulation for about 100 times, the average value as our results. From the graph, we know that the spreading speed in Case 1 is similar to that in the standard SI model until the time t_D , after some efficient measure taken, the number of increased infected ones at every simulation time decrease sharply. So is that in Case 2. The maximum of increased infected ones in Case 2 is less than other models because of the less mean infection rates.

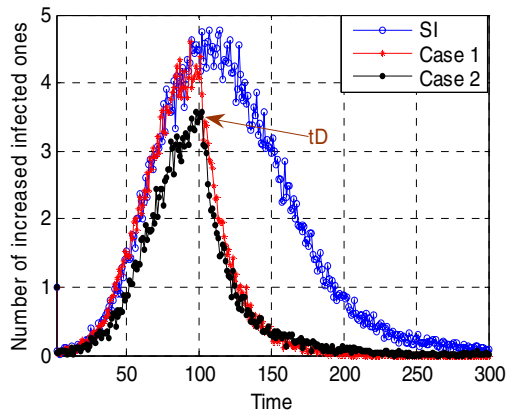


Fig. 3. Average rate of increased infected individuals in dependence of the simulation time for SI model, Case 1, Case 2. The infection rate in Standard SI model, mean infection rate in Case 1 and initial infection rate in Case 2 are all 0.01, $t_D=100$, other parameters settings are introduced in section 2. We run our simulation for about 100 times, the average value as the result.

Second, we study the steady infected rates for different parameter t_D , which were shown in the fig.4, We run our simulation for about 100 times, the average rates of final infected ones as results. From the graph, we can see with the increase of t_D , more people will be infected. And the rates increased at a exponential form when the parameter t_D is less than 150, so it's better to adopt the preventive and control measures as soon as possible. Another interesting phenomenon is the rates of final infected ones in Case 2 is larger than that in Case 1 when the parameter t_D is small, this is because, the parameter settings cause the preventive and control measures in Case 2 less effective than that in Case 1, but when t_D is large enough, the process in Case 2 before the time t_D is slower than that in Case 1, although the preventive and control measures in Case 2 is less effective, the rate of final infected ones are still lower than that in Case 1.

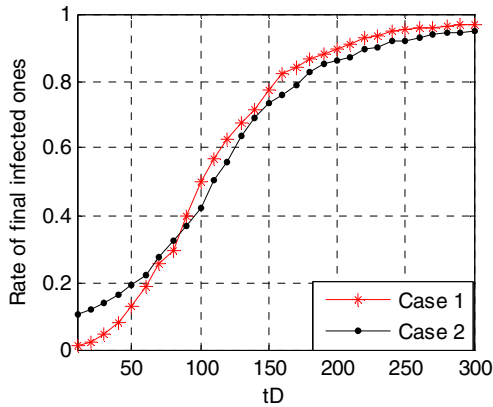


Fig. 4. Average rate of final infected individuals in dependence of the parameter t_D for Case 1, Case 2. The mean infection rate in Case 1 and initial infection rate in Case 2 are both 0.01, other parameters settings are introduced in section 2. We run our simulation for about 100 times, the average value as the result.

(3) Distribution of final infection rates

In Case 1, the infection rates are assumed to be normal distribution, but after the time t_D , preventive and control measures are taken, so susceptible individual's infection rates will decrease, how's the final infection rates? That's what we should study in this section. By means of simulation, the distribution of the final infection rates can be seen in fig.5, most of individuals has almost zero infection rates, that means many people won't be infected in the final steady state, as can also be seen in fig.2, because after a period of time, people gain more information about this epidemic, and some efficient measures will bring into effects, finally the infection rates will tend to be zeros.

However, in Case 2, we can see obvious bimodal distribution, this difference can be explained as follow: at the initial time of epidemic break, few knowledge about the epidemic, people pay no attention, and then many people will be infected, while people realize the severity of this epidemic, many people will take many efficient measures, and a large portion of people will not be infected in the end.

This kind of phenomenon also exists in Case 2, but as the infection rates were known beforehand, and they are distributed inside a range. So the bimodal phenomenon is not obvious in Case 1, but we still can see a little peak near the 0.01.

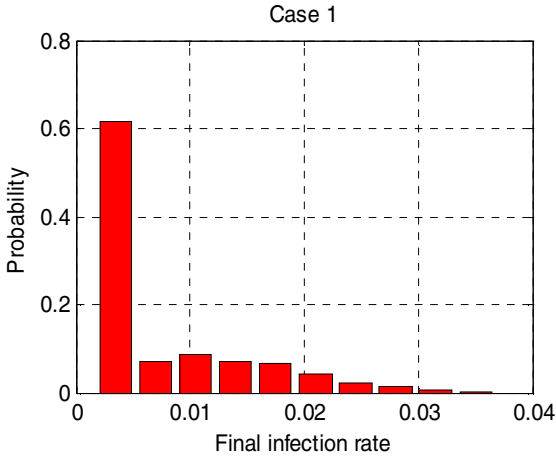


Fig. 5. Distribution of individual's final infection rates for different m in Case 2

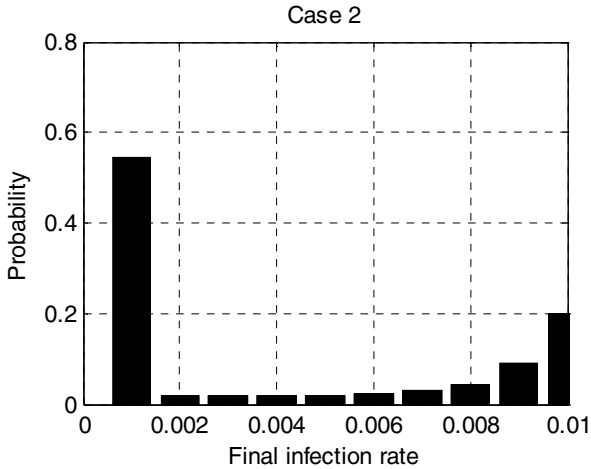


Fig. 6. Distribution of individual's final infection rates for Case 1, Case 2. The mean infection rate in Case 1 and initial infection rate in Case 2 are both 0.01, $t_D=100$, other parameters settings are introduced in section 2. We run our simulation for about 100 times, the average value as the result.

5 Conclusion and Discussion

Almost all the previous studies about the epidemic spreading models in scale-free networks essentially assume that the infection rates for all individuals are the same.

In real epidemic events, individual's infection rate are dominant by his resistance to diseases and preventive and control measures, special body mass might have better resistance to a kind of epidemic, and better preventive and control measures bring lower probability to be infected. To simulate the epidemic process, it's essential to consider the individual's different infection rate. In our paper, we proposed two kinds of SI models with variant infection rates. In Case 1, we know every individual's infection rate, satisfying a kind of distribution. We use several random equations to describe the epidemic processes equally, and from the theory, we know that if the mean infection rate equals the infection rate in standard SI model, the epidemic processes before the time t_D in two models will be almost the same, the simulations testify the results. In Case 2, we don't know every body's resistance to the diseases, considering that a susceptible individual will have a lower infection rate if he is not infected after limited number of contacts with infected individuals. The epidemic process is relatively slow because of the decreased mean infection rates.

Moreover, we know that the parameter t_D has a large impact on the epidemic process, after the time t_D , the epidemic process become slow, and the number of increased infected ones at every simulation time decrease sharply, while the mean infection rates in both Case 1 and Case 2 will decrease, finally many people won't be infected, corresponding to almost zero infection rate, this can explain why most people will hardly be infected in most of epidemic break. What's more, from the distribution of final infected rates, we know that the distribution in Case 2 is bimodal, correspond to a real life to a society in which a relative large portion of people will be infected because of less resistance against the epidemic while another large portion of people will hardly be infected in the end due to efficient preventive and control measures.

We put forward a SI model with variant infection rates to describe the epidemic process, give us some new understanding about epidemic process. But some parameter settings are artificial, such as m_1 , m_2 , k_1 , k_2 , and so on; this parameters may be different for different epidemic events, so it should be more appropriate to combine the real epidemic break to affirm these parameters. Whether the thought on other models and networks can get some useful results is worthy of further study.

Acknowledgments. The authors thank Professor Qi Fei, Yi Shen and Zhigang Zeng for their valuable comments and suggestions. This work has been partly supported by the Natural Science Foundation of China under Grant No. 60773188 and China Postdoctoral Science Foundation (CPSF Grant 20080430961).

References

1. Pastor-Satorras, R., Vespignani, A.: Epidemics and Immunization in Scale-free Networks. In: Bornholdt, S., Schuster, H.G. (eds.) Handbook of Graph and Networks. Wiley-VCH, Berlin (2003)
2. Zhou, T., Fu, Z.Q., Wang, B.H.: Prog. Nat. Sci. 16, 452 (2006)
3. Pastor-Satorras, R., Vespignani, A.: Phys. Rev. Lett. 86, 3200 (2001)
4. Pastor-Satorras, R., Vespignani, A.: Phys. Rev. E 63, 066117 (2001)
5. May, R.M., Lloyd, A.L.: Phys. Rev. E 64, 066112 (2001)
6. Moreno, Y., Pastor-Satorras, R., Vespignani, A.: Eur. Phys. J. B 26, 521 (2002)

7. Ke, H., Yi, T.: Temporal Behaviors of Epidemic Spreading on the Scale-free Network. *Physica A* 373, 845–850 (2007)
8. Zhou, et al.: Behaviors of Susceptible-infected Epidemics on Scale-free Networks with Identical Infectivity. *Physical Review* 74, 056109 (2006)
9. Barthélemy, M., Barrat, A., Pastor-Satorras, R., Vespignani, A.: *Phys. Rev. Lett.* 92, 178701 (2004)
10. Barthélemy, M., Barrat, A., Pastor-Satorras, R., Vespignani, A.: *J. Theor. Biol.* 235, 275 (2005)
11. Zhou, T., Yan, G., Wang, B.H.: *Phys. Rev. E* 71, 046141 (2005)
12. Olinky, R., Stone, L.: *Phys. Rev. E* 70, 030902 (2004)
13. <http://www.zyi.ks.edu.tw/data/anisars.doc>
14. Fa, D.S., Zhai, F.Y., Ge, K.Y., Chen, F.N.: Distributions of Body Mass Index of Chinese Adults. *Journal of Hygiene Research* 6, 209–215 (2001)
15. Zhen, Z.K., Ma, L., Chen, H.Y.: The Study of the Distribution of the Chinese BMI. *Mathematical Theory and Application* 3, 312–315 (2000)

Interdependency Analysis of Infrastructures

Zijun Mao^{1,2}, Liu Hong^{1,*}, Qi Fei¹, and Ming OuYang¹

¹ Institute of System Engineering, Huazhong University of Science and Technology, Wuhan 430074, China

² Key Lab. for Image Processing & Intelligent control, Huazhong University of Science and Technology, Wuhan 430074, China

Abstract. Electric power, potable water, telecommunications, natural gas, and transportation are examples of critical infrastructures, the intrinsic feature of which are suitable for network analysis. This paper proposes a method framework based on the complex network to explore the interdependency of these infrastructures. We describe the topological characterization of two interdependent small-sized real networks, based on which the interdependent model is devised. We define the interdependent factor to characterize the direct interdependency of two infrastructures, and also its fluctuation in the situations of random removal and targeted removal on the electric and water infrastructures is studied. Furthermore we propose a matrix M to reflect all direct and indirect interactions among infrastructures, which also makes contributions to the understanding of infrastructures catastrophes.

Keywords: Interdependent systems, Complex networks, Critical infrastructure protection.

1 Introduction

The importance of cross-sector infrastructure interdependencies was first highlighted at the national level in 1997 when the President's Commission on Critical Infrastructure Protection released their landmark report *Critical Foundations: Protecting America's Infrastructures* [1]. The report emphasized that the security, economic prosperity, and social well being of the nation were dependent on the reliable functioning of our increasingly complex and interdependent infrastructures.

In defining their case for action, the Commission noted that the energy (which they referred to as the lifeblood of our interdependent infrastructures) and communications infrastructures created an increased possibility that a rather minor and routine disturbance can cascade into a regional outage. They further concluded that the technical complexity of the infrastructure might also lead to the neglect of interdependencies and vulnerabilities until a major failure occurred. The August 14, 2003, blackout that large portions of the Midwest and Northeast United States and Ontario, Canada, experienced an electric power outage dramatically left us an example of the highly

* Corresponding author.

complex technical challenge that we may face in preventing cascading impacts [2]. Therefore there is an urgent need to better understand and develop better idea to handle problems related to the modeling complexity and the interconnected large scale complex critical infrastructures.

Two popular approaches for reliability analysis in interdependent infrastructures are agent-based simulation and input-output analysis. The core idea behind the development of agent-based simulations for this application is that individual components and subsystems can be represented as agents which are designed to evolve and interact with each other, then emergent behaviors (i.e. interdependencies) can be identified on them (e.g. [3], [4], [5] and [6]). Agent-based simulation is also being used to investigate the electric power and natural gas markets (e.g. [7], [8], [9], and [10]). The main idea of market-level models is to trace behaviors of economic agents in these industries.

Input-output analysis has traditionally been used to model the interactions among sectors of the economy and forecast the impacts that the changes in one part of the economy may have on the performance of the others. Haimes and Jiang [11] suggest that the same modeling paradigm may be useful to model the interactions and interdependencies within and across infrastructures. While many of the agent based simulations for reliability analysis contain very detailed system representation, the input-output modeling is likely to be very aggregate.

We notice that characterizing the interdependency among infrastructures is still a major challenge in the recent study. This is because critical infrastructures interdependency may be characterized and measured from a multitude of perspectives and attributes. In this paper by taking into account the large number of interconnected subsystems and nodes, we propose a method to explore the interdependent response of the infrastructures from the perspective of the system's complexity.

In the next section, we present the topological characterization of two interdependent small-sized real networks, based on which the interdependent model is devised. Section 3 is intended to illustrate the interdependent response of two infrastructure networks, where we give the definition of interdependent factor. In section 4 we propose a matrix M to reflect all direct and indirect interactions among infrastructures. Section 5 summarizes our results and surveys the potential improvement in the future.

2 Interdependent Infrastructure Network Model

Our critical infrastructures, such as telecommunications, electrical power systems, gas and oil storage, transportation, water-supply systems and emergency services, are composed of many functional subsystems and the connections between them. By taking the subsystems as nodes and the connections as links, we can use network-graph to characterize the topology of infrastructures, and the interdependencies among different infrastructures are interpreted as links between interdependent infrastructure nodes.

2.1 Network Described Parameters

A network or graph G is a set of elements referred to as vertices (nodes) with connections between them denoted by edges (links). $G = g(V, E)$ is employed to represent

the electric power network. The vertices of a graph G are defined as the vertex set of G , denoted by $V(G)$, and the edges are defined as the edge list of G , denoted by $E(G)$. We analyze several main parameters of infrastructure network, including: degree distribution, average path length, network efficiency and clustering coefficient. These parameters are defined as follows:

a) Vertex Degree and Its Distribution

The vertex degree, $d(v)$, of undirected graphs is the number of edges connected to a vertex, v . The average vertex degree, $d(G)$, is the average of all $d(v)$ for $v \in V(G)$.

b) Average Path Length

The characteristic average path length, L , of a graph is the mean of the shortest path lengths, $d(v_i, v_j)$, connecting each vertex $v \in V(G)$ to all other vertices. This parameter can be regarded as a global indicator of network connectivity. For general undirected graphs this parameter is calculated as follow:

$$L = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \neq j} d(v_i, v_j).$$

If L is large, the dynamics within the network are slow due to the many intermediate steps to connect any two nodes.

c) Network Efficiency

In the event that two vertices are not connected at all, or become disconnected due to disruption, their shortest path length d becomes infinity. One way to handle infinite values is to calculate network efficiency, E , it is calculated as follow:

$$E = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \neq j} \frac{1}{d(v_i, v_j)}.$$

d) Clustering Coefficient

An important concept in networks, referred to as neighborhood, is critical for the calculation of clustering coefficients. Clustering coefficient is the probability of one node linked by any two neighbor nodes. The network average clustering coefficient is the mean of all nodes' clustering coefficients. For node i , G_i is its neighbor nodes, k_i is its vertex degree, $|G_i|$ is the number of nodes in G_i connected to i , so clustering coefficient C_i and network average clustering coefficient $\langle C \rangle$ are calculated as:

$$C_i = \frac{|G_i|}{\frac{1}{2}k_i(k_i-1)}, \quad C = \sum_i \frac{C_i}{N}.$$

2.2 Infrastructure Networks

In this subsection, two small-sized simplified critical infrastructures are chosen to estimate their topological properties and model the interdependency. The systems are corresponding to a sample of the electric power and water distribution networks of a major city in Central China. Network elements are gate stations, electric substations, and transmission lines for the power grid. Storage tanks, pump stations, and distribution pipelines constitute the elements of the water network.

Fig. 1 displays the topology of these selected networks and identifies the role of the elements. The topological properties of these networks are summarized in Table 1.

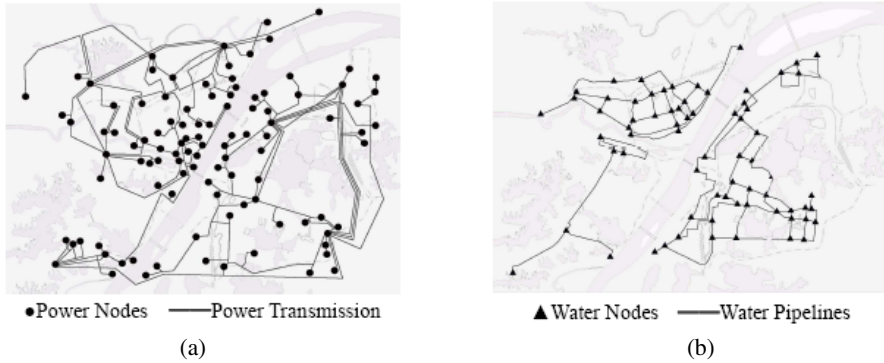


Fig. 1. (a) Network topology of simplified electric power network; (b) network topology of simplified water distribution network

Table 1. Fundamental Topological Properties of power network and water network

Network type	Vertex degree	Clustering coefficient	Path length
Power network	3.73	0.451	11.7
Water network	3.14	0.174	6.08

Dependence of the water distribution network on the electric power grid is illustrated by its demand of power for appropriate functioning of pump stations, lift stations, and control units. Considering each power node serves water node of its neighboring area, we get the interdependent matrix $\mathbf{I}(i,j)$ and each nonzero element of row i and line j in \mathbf{I} indicates the i th element of the water network that depends on the j th element of the power network.

3 Interdependent Responses of Networks

In this section, we will utilize computer simulation to analyze the interdependent response of infrastructure networks to disruption. Any failure in a power grid element will disconnect the water elements that interact with it, if they exist, and eliminate additional water elements if they belong to an extended power service area that loses its power supply, this is showed in Fig.2.

In simulation, first, we fix the fraction of removed power grid elements, which is the same with the water network, and remove one node of each network at a time until the specified removal fraction level is reached. Then, subject both networks to simultaneous failure under random removal and targeted removal. Random removal is interpreted as pure random failure of nodes, and targeted removal is interpreted as targeted removal of nodes based upon their initial vertex degree distribution. The removal

process is illustrated as follows: choose a removal strategy (random removal or targeted removal), and then randomly sample a node to remove using the chosen strategy, repeat this procedure until the network is depleted of its nodes. Network response is measured by the loss of global connectivity as captured by its connectivity efficiency, E .

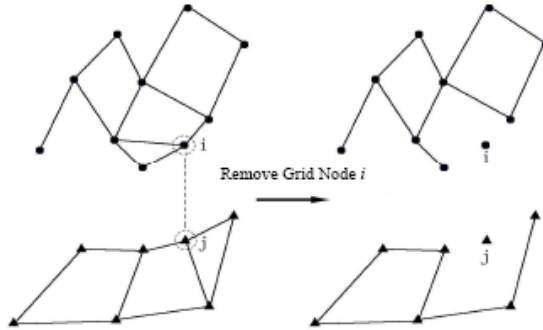


Fig. 2. Interdependent Node Removal Example

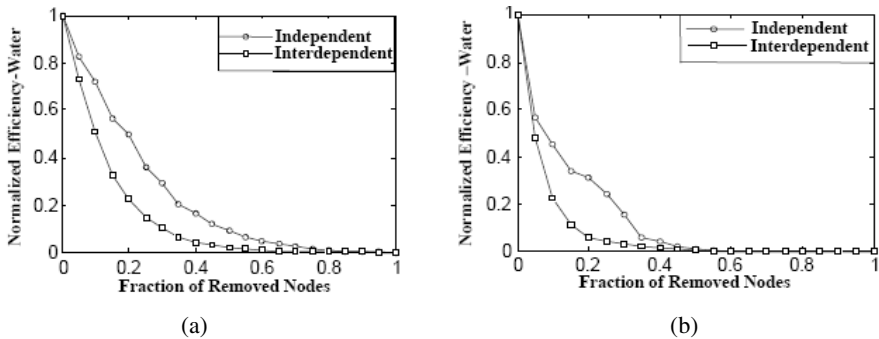


Fig. 3. Shows the impact of the strength of coupling on water network performance when the two networks are simultaneously subjected to (a) random removal and (b) targeted removal, and the responses are measured by water network efficiency E

Fig. 3 suggests that the targeted removal proves to be more harmful than random removal. Also, water network makes better response to random selection of nodes than to targeted selection. This result is expected because there are some nodes with a vertex degree larger than the average. These vertices with larger-than-average degree are more difficult to be chosen by uniform sampling. No matter it is random or targeted removal, the interdependency worsens the response of the water network.

Additional information regarding the way in which these interdependencies manifest is provided by a parameter introduced as the “interdependent factor,” f . This parameter is defined as the absolute difference between the independent and interdependent responses, which is normalized by the maximum independent response attained at any removal fraction. If E_1 , is denoted as the independent response

efficiency of water network and E_2 is denoted as the interdependent response efficiency, f is calculated as $f = \frac{|E_2 - E_1|}{\max(E_1)}$.

Fig. 4 illustrates the interdependent factor's fluctuation effect, whose effect on the water system response is shown as the power grid undergoing disruptions according to two removal strategies (random removal, and targeted removal).

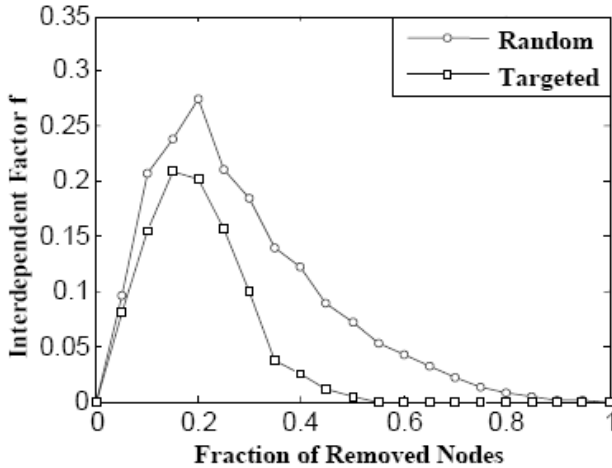


Fig. 4. Interdependent factor's fluctuation under random and targeted removal strategies

Smaller interdependent factor is observed when both networks are simultaneously disrupted by targeting their most connected nodes. Picking the most connected nodes of the power network, instead of picking them at random, appears to cause no significant interdependent effects on the water network. If the power network is being severely disrupted, then why does the water network remain invariant? The reason is that the most connected nodes of the power grid are not necessarily the nodes that facilitate the interaction with other systems. The water network depends on power nodes that are not the most connected ones.

4 Assessment of Direct and Indirect Effects of Interdependent Networks

In the above section, we have analyzed the interdependency of two infrastructure networks and give the interdependent factor, f , to characterize the interdependency. Based on this definition it is feasible to determine the direct interdependency f_{ij} of one infrastructure network j on another one i , and we can summarize the direct interdependency of infrastructures by a matrix $\mathbf{F}=\{ f_{ij} \}$.

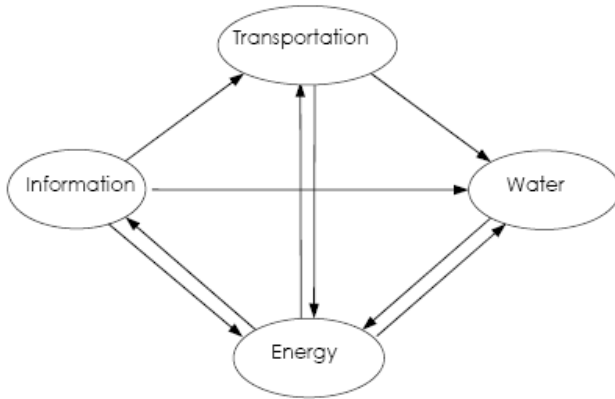


Fig. 5. Interdependent Infrastructure Networks Example

Usually, during a disaster the responsibilities have only dissatisfactory information and short time to decide, so in many cases we will take into account only direct impacts on other infrastructures. In the worst case, this may lead to the opposite other than the desired result, if the feedback effects exceed the direct influence. Therefore, it would be better to know the implications on the whole system, and it is necessary to take into account all factors which are relevant during the catastrophe.

So we represent both direct and indirect effects of infrastructure j on infrastructure i by M_{ij} , and summarize all these influences by a matrix $\mathbf{M}=(M_{ij})$, which is determined from the matrix \mathbf{F} of direct interdependency. We use a formula as $\mathbf{M}=\mathbf{F}^k$. The expression \mathbf{F}^k reflects all influences over $k - 1$ nodes and k links, i.e., $k = 1$ corresponds to direct influences, $k = 2$ to feedback loops with one intermediate node, $k = 3$ to feedback loops with two intermediate nodes, etc. The matrix $\mathbf{M}=(M_{ij})$ could summarize all direct influences and feedback effects among infrastructures, and also can be used to estimate the temporary development of catastrophes among infrastructures.

5 Conclusion

The networked nature of infrastructure systems enables the objective representation of them by using tools from graph theory. In this work, the topological properties of two real networks are firstly illustrated. Electric power and water distribution can be characterized in terms of their global and local connectivity, their vertex degree distribution, average path length, and clustering coefficient. Then we introduce a procedure to explore the interdependencies between these two real infrastructure networks, and we formulate interdependent factor to characterize the interdependency.

To better understand network failure mechanisms it is necessary to further analyze multiple infrastructure systems as a single interacting entity. Therefore we propose a matrix \mathbf{M} to represent all direct and indirect interactions among infrastructures, and \mathbf{M} also makes contributions to the understanding of infrastructures catastrophes.

Interdependent infrastructure analysis can enhance loss estimation methodologies and suggest strategies for robust design and growth of infrastructures. Investors, owners, and operators of utility companies can use the results from an interdependent analysis to make better decisions on prioritizing scarce resources for mitigation actions.

Acknowledgments. This work has been partly supported by the Natural Science Foundation of China under Grant No. 60773188 and China Postdoctoral Science Foundation (CPSF Grant 20080430961).

References

1. President's Commission on Critical Infrastructure Protection, *Critical Foundations: Protecting America's Infrastructures* (1997), <http://www.fas.org/sgp/library/pccip.pdf>
2. U.S.-Canada Power System Outage Task Force, *Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations* (2004)
3. Tomita, Y., Fukui, C., Kudo, H.: Cooperative Protection System with an Agent Model. *IEEE Transactions on Power Delivery* 13, 1060–1066 (1998)
4. Wildberger, A.M.: Modeling with Independent Intelligent Agents for Distributed Control of the Electric Power Grid. In: *59th Annual American Power Conference*, Chicago, pp. 361–364 (1997)
5. Wildberger, A.M.: Modeling the Infrastructure Industries as Complex Adaptive Systems. In: *Simulation International XV*, San Diego, CA, pp. 168–173 (1998)
6. Amin, M.: Toward Secure and Resilient Interdependent Infrastructures. *Journal of Infrastructure Systems* 8, 67–75 (2000)
7. North, M.: SMART II: the Spot Market Agent Research Tool version 2.0 plus Natural Gas. Argonne National Laboratory, Illinois (2000)
8. North, M.: SMART II+: the Spot Market Agent Research Tool version 2.0. Argonne National Laboratory, Illinois (2000)
9. Tsoukalas, L.H., Uluyol, O.: Anticipatory Agents for the Deregulated Electric Power System. In: *Workshop on Agent Simulation: Applications, Models, and Tools*, pp. 114–123. The University of Chicago, Chicago (1999)
10. Thomas, R., Mount, T.D., Zimmerman, R.: Testing the Effects of Price Responsive Demand on Uniform Price and Soft-Cap Electricity Auctions. In: *35th Hawaii International Conference on Systems Sciences*, pp. 121–130. IEEE Press, New York (2002)
11. Haimes, Y., Jiang, P.: Leontief-Based Model of Risk in Complex Interconnected Infrastructures. *Journal of Infrastructure Systems* 7, 1–12 (2001)

Back Propagation Neural Network Based Lifetime Analysis of Wireless Sensor Network

Wenjun Yang, Bingwen Wang, Zhuo Liu, and Xiaoya Hu

Dept. of Control Science and Engineering
Huazhong Univ. of Science and Technology
Wuhan, 430074, China
yangwjn@163.com

Abstract. As large amount of energy constrained nodes are randomly distributed in network, the entire lifetime of wireless sensor network is difficult to estimate. In this paper, a back propagation (BP) neural network based Markov model is presented to calculate the lifetime of wireless sensor network. BP neural network is employed to reduce the calculation difficulty of Markov state equation. The simulation results indicate that this method gives the value of maximum lifetime exactly and its computing complexity is low. The quantitative degree of computing reliability relative error between the results of three layer BP neural network and the Markov model is 10^{-4} .

Keywords: Wireless sensor networks, Network lifetime analysis, BP neural network, Markov.

1 Introduction

The wireless sensor network (WSN) consists of a large number of smart tiny sensor nodes. Each individual node in the network can monitor its local region and communicate with other nodes through a wireless channel. The lifetime of WSN is defined as the time interval from the initial state of network to the failure state in which the network can not collaboratively produce a high-level representation of the environment's states. Energy saving which will prolong the lifetime of network is one of the central issues in research field of WSN, as the energy of node is limited.

A few progresses about maximum lifetime of WSN have been obtained in some research results, which reveal the bound or expectation value of the maximum lifetime. The upper bound on the lifetime of sensor networks is first put forward in [1], and then [2] introduces a data fusion algorithm based on [1] to calculate the bound through optimal routing and role assignment algorithm. A mathematical analysis for lifetime by modeling data-generation process is provided by [3]. The formalization of the lifetime-maximizing problem to a multi-source multi-sink flow-maximizing problem on a directed graph with arc and vertex capacity powers is presented in [4]. Meanwhile, [5] demonstrates that a type-2 fuzzy membership function is most appropriate to model a single node lifetime. The lifetime-maximizing problem can also be formalized by linear programming which needs related heuristic algorithms to calculate the approximate solutions [6-10]. In this paper, we address the issue of lifetime analysis and estimation

for wireless sensor networks in which the sensor nodes are deployed at desired locations. Our approach is entirely different from all prior research. Instead of trying out various probability basis, we propose to apply an BP neural network to present a Markov model for lifetime analysis and estimation in a wireless sensor network.

The rest of this paper is organized as follows. In section 2, we detail the reliability model of discrete state Markov chain. Section 3 gives an overview of neural network based Markov model. In section 4, we introduce the state description of distributed wireless sensor network and apply them to a Markov model for lifetime estimating. Simulation results and discussions are presented in section 5. Conclusion remarks are collected in section 6.

2 Reliability Model of Discrete State Markov Chain

The discrete Markov model including n states is shown in Fig.1. Each state defines the physical state of the system in a certain time. When a fault occurs at runtime, state of system may be changed from one to another. The state 1 is the initial state and the system works well. The final one which is denoted by state n means the system is failure completely. The directed connections between two states represent the transition probabilities of different states, implicitly expressing the processes of system upgrade or degradation. a_{ij} is the transition probability from state i to state j , as depicted in Fig.1.

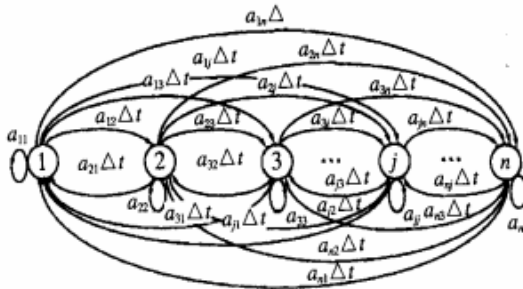


Fig. 1. Discrete Markov model including n states

The state equations of discrete Markov model can be expressed as follows:

$$P(t + \Delta t) = [P_1(t + \Delta t), P_2(t + \Delta t), \dots, P_j(t + \Delta t), \dots, P_n(t + \Delta t)] \tag{1}$$

$$P(t) = [P_1(t), P_2(t), \dots, P_j(t), \dots, P_n(t)] \tag{2}$$

$$B = \begin{pmatrix} a_{11} & a_{21}\Delta t & a_{31}\Delta t & \dots & a_{j1}\Delta t & \dots & a_{n1}\Delta t \\ a_{12}\Delta t & a_{22} & a_{32}\Delta t & \dots & a_{j2}\Delta t & \dots & a_{n2}\Delta t \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ a_{13}\Delta t & a_{2j}\Delta t & a_{3j}\Delta t & \dots & a_{jj} & \dots & a_{nj}\Delta t \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ a_{1n}\Delta t & a_{2n}\Delta t & a_{3n}\Delta t & \dots & a_{jn}\Delta t & \dots & a_{nn} \end{pmatrix}, a_{ji} = 1 - \sum_{k=1(k \neq j)}^n a_{jk}\Delta t \tag{3}$$

$$P(t + \Delta t) = BP(t) , \tag{4}$$

where $P_j(t)$ is the probability of the system which is in state of i at the time (t) . The element a_{ij} in transition matrix B represents the transition probability from state i to state j .

If the fault states include the states from j to n , the reliability of system is:

$$R(t) = 1 - \sum_{i=j}^n P_i(t) . \tag{5}$$

3 Markov Model Based on Neural Network

The architecture of the neural network model proposed here is shown in Fig.2. It can be seen that there is an input layer, a hidden layer and an output layer in the neural network. Each layer consists of n neurons which equals the count of states in Markov model. V_{nj} represents the connection weight between the output of the i th neuron in input layer and the input of the n th neuron in hidden layer. The connection weight between the output of the j th neuron in hidden layer and the input of the n th neuron in output layer is represented by W_{nj} . The relations among the input layer, hidden layer and output layer of neural network at the running time are summarized as follows:

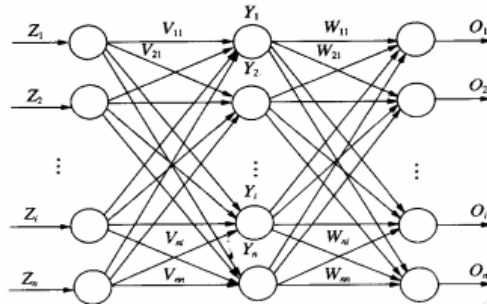


Fig. 2. Neural network model including 3 layers

$$Z = [Z_1, Z_2, \dots, Z_n]^T = P(t) \tag{6}$$

$$Y = [Y_1, Y_2, \dots, Y_n]^T = P(t + \Delta t) \tag{7}$$

$$O = [O_1, O_2, \dots, O_n]^T = P(t + 2\Delta t) . \tag{8}$$

The matrixes of weight value are expressed as follows:

$$V = \begin{pmatrix} V_{11} & V_{12} & \dots & V_{1n} \\ V_{21} & V_{22} & \dots & V_{23} \\ \dots & \dots & \dots & \dots \\ V_{n1} & V_{n2} & \dots & V_{nn} \end{pmatrix} , \quad W = \begin{pmatrix} W_{11} & W_{12} & \dots & W_{1n} \\ W_{21} & W_{22} & \dots & W_{23} \\ \dots & \dots & \dots & \dots \\ W_{n1} & W_{n2} & \dots & W_{nn} \end{pmatrix} . \tag{9}$$

In addition, the output vector of expectation is defines as:

$$D = [D_1, D_2, \dots, D_n]^T \tag{10}$$

The state of system is normal in initial stage, so the initial inputs of neural network are $[Z_1, Z_2, \dots, Z_n] = [1, 0, \dots, 0]$.

The following equations can be obtained from the analysis of the relations among the three layers of neural network in Fig.2:

$$Y = VZ, \quad O = WY \tag{11}$$

The energy function of neural network can be expressed as:

$$E = \sum_{k=1}^n (O_k - D_k)^2 \tag{12}$$

Based on the learning and self-adaptive function of neural network, the connection weight can be adjusted properly to converge E to the preset value. When E converges to the preset value, related reliability parameters which meet the design requirements can be calculated through formula (13). In order to accelerate the convergence rate of neural network, we use gradient descending rules to modify the connection weight W_{ij} .

O_k is the actual output of neurons k , corresponding to the probability of system in state k . Meanwhile, D_k is the expectation output of neurons k in output layer, corresponding to design goal of system in state k .

And then, weight adjustment quantity between hidden layer and output layer is obtained through BP algorithm:

$$\Delta W_{ji} = -2\eta \sum_{k=1}^n \delta_{ok} \frac{\partial O_k}{\partial W_{ji}}, \quad \delta_{ok} = (O_k - D_k), k = 1, 2, \dots, n \tag{13}$$

where δ_{ok} is the error value, produced by neurons k in output layer, η is learning factor.

The weight adjustment quantity between input layer and hidden layer is:

$$\Delta V_{ji} = -2\eta \sum_{k=1}^n \delta_{yk} \frac{\partial E}{\partial V_{ji}}, \quad \delta_{yk} = \sum_{p=1}^n \delta_{op} \frac{\partial Q_p}{\partial Y_k}, \quad k = 1, 2, \dots, n \tag{14}$$

where δ_{yk} is the error value, produced by neurons k in hidden layer.

The expectation reliability of system is the output of the neural network. Related design parameters including reliability $R(t)$, failure rate f and repair rate v can be obtained through the weight of neural network.

4 Markov Model of the Wireless Sensor Network

The system has two kinds of failures, the first one is that some parts of the network are failed to monitor data, and the second one is the failure of whole network. dt expresses the length of the time during state changing.

4.1 State Descriptions of the Wireless Sensor Network

The wireless sensor network consists of five states, defined as follows:

- State 0:** system is in a normal state;
- State 1:** some of nodes in the system are sleeping to save energy;
- State 2:** some of nodes fail to monitor data, but vicinal region can be monitored by other nodes;
- State 3:** the nodes in a region fail completely, resulting in inefficiency of monitoring in the region;
- State 4:** all of the nodes in network fail completely.

In these states, state 0 is the initial state in which system works well. State 5 is the final state of the system which is completely failed in this state. The directed connection between two different states signifies the transition probability, which represents the physical process of system at running time. The Markov model of wireless sensor network is shown in Fig 3.

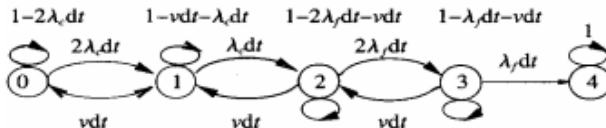


Fig. 3. Markov model of wireless sensor network

The formulas of Markov model of wireless sensor network are shown as follows:

$$\begin{cases}
 P_0(t + \Delta t) = (1 - 2\lambda_c \Delta t)P_0(t) + v \Delta t P_1(t) \\
 P_1(t + \Delta t) = 2\lambda_c \Delta t P_0(t) + (1 - v \Delta t - \lambda_c \Delta t)P_1(t) + v \Delta t P_2(t) \\
 P_2(t + \Delta t) = \lambda_c \Delta t P_1(t) + (1 - 2\lambda_f \Delta t - v \Delta t)P_2(t) + v \Delta t P_3(t) \\
 P_3(t + \Delta t) = 2\lambda_f \Delta t P_2(t) + (1 - \lambda_c \Delta t - v \Delta t)P_3(t) \\
 P_4(t + \Delta t) = \lambda_f \Delta t P_3(t) + P_4(t)
 \end{cases} \quad (15)$$

The reliability of the system is:

$$R(t) = P_0(t) + P_1(t) + P_2(t) + P_3(t) .$$

At the initial stage, the system is working with non-failure, so the probability of each state is:

$$P_0(0) = 1, P_1(0) = P_2(0) = P_3(0) = P_4(0) = 0 .$$

4.2 Neural Network Based Markov Model of Wireless Sensor Network

As shown in Fig. 4, the neural network based Markov model of wireless sensor network has five neurons in each layer. The count of neurons is the same as the number of states in Markov model of system. At any specific time (t) when system is running, the input

vector(Z), output vector of hidden layer and output layer(O), the increment of weight ΔW_{ji} and ΔV_{ji} are:

$$\begin{cases} Z_1 = P_0(t) \\ Z_2 = P_1(t) \\ Z_3 = P_2(t) \\ Z_4 = P_3(t) \\ Z_5 = P_4(t) \end{cases} \begin{cases} Y_1 = P_0(t + \Delta t) \\ Y_2 = P_1(t + \Delta t) \\ Y_3 = P_2(t + \Delta t) \\ Y_4 = P_3(t + \Delta t) \\ Y_5 = P_4(t + \Delta t) \end{cases} \begin{cases} O_1 = P_0(t + 2\Delta t) \\ O_2 = P_1(t + 2\Delta t) \\ O_3 = P_2(t + 2\Delta t) \\ O_4 = P_3(t + 2\Delta t) \\ O_5 = P_4(t + 2\Delta t) \end{cases}$$

where the initial state of neural network is: $Z_1 = 1, Z_2 = Z_3 = Z_4 = Z_5 = 0$.

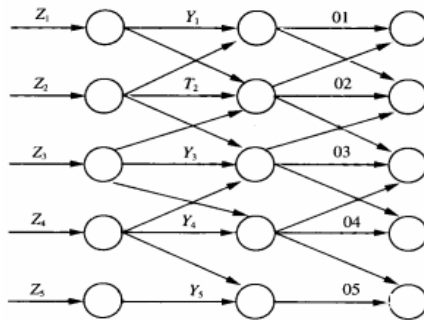


Fig. 4. Neural network based Markov model of wireless sensor network

5 Simulation Results and Analysis

With different initial parameters of λ_c, λ_f, v and $R(t)=0.8800$, the reliability $R(t)$ of convergent point is 0.880130 and precision is 0.005 after 22 times iteration. Related parameters can also be calculated by the connection weights, and $\lambda_c=0.0019, \lambda_f=0.00012$ and $v=0.00336$. So the lifetime of system is:

$$LT = \frac{1}{\lambda_c + \lambda_f} \approx 471h .$$

The reliability degree is 0.8806, calculated by Markov model directly. The quantitative degree of computing reliability relative error between the three layer BP neural network and the Markov model is 10^{-4} . So it is practicable to calculate the lifetime of wireless sensor network.

6 Conclusions

It is significant to research the calculation of WSN lifetime in both theory and practical. Calculating the lifetime of wireless sensor network is a valuable issue with theoretical and actual significance. In this paper, a back propagation neural network based Markov

model is presented to calculate the lifetime. Not only is the exactly maximum lifetime given by this solution, but also has low complexity and fast convergence. Indication by simulation, the algorithm of neural network is suitable for calculating the maximum lifetime of wireless sensor network. We believe that our approach opens up a new vision for research on wireless sensor network lifetime analysis.

The future work will focus on lifetime evaluation under the circumstances that the task scheduling is variable and how the estimated network lifetime could be used to accommodate the scheduling change.

Acknowledgments. The authors are grateful for the financial support provided by the National Nature Science Foundation of China (No. 60773190 and 60802002).

References

1. Bhardwaj, M., Chandrakasan, A., Garnett, T.: Upper Bounds on the Lifetime of Sensor Networks. In: IEEE International Conference on Communications, pp. 785–790. IEEE Press, Helsinki (2001)
2. Bhardwaj, M., Chandrakasan, A.: Bounding the Lifetime of Sensor Networks via Optimal Role Assignments. In: IEEE INFOCOM, pp. 1587–1596. IEEE Press, New York (2002)
3. Rai, V., Mahapatra, R.N.: Lifetime Modeling of a Sensor Network. In: Design, Automation and Test in Europe, Munich, pp. 1530–1591 (2005)
4. Yantao, P., Peng, W., Xicheng, L.: Maximum Flow Based Model and Method of the Maximum Lifetime Problem of Sensor Networks. In: The Sixth World Congress on Intelligent Control and Automation, vol. 1, pp. 3623–3626. IEEE Press, New York (2006)
5. Haining, S., Qilian, L., Jean, G.: Wireless Sensor Network Lifetime Analysis Using Interval Type-2 Fuzzy Logic Systems. *J. Fuzzy Systems* 16, 416–427 (2008)
6. Chang, J.H., Tassiulas, L.: Energy Conserving Routing in Wireless Ad-hoc Networks. In: IEEE INFOCOM 2000, pp. 22–31. IEEE Press, New York (2000)
7. Chang, J.H., Tassiulas, L.: Fast Approximation Algorithms for Maximum Lifetime Routing in Wireless Ad-hoc Networks. In: Pujolle, G., Perros, H.G., Fdida, S., Körner, U., Stavrakakis, I. (eds.) NETWORKING 2000. LNCS, vol. 1815, pp. 702–713. Springer, Heidelberg (2000)
8. Dasgupta, K., Kalpakis, K., Namjoshi, P.: Efficient Algorithms for Maximum Lifetime Data Gathering and Aggregation in Wireless Sensor Networks. *J. Computer Networks* 42, 697–716 (2003)
9. Madan, R., Lall, S.: Distributed Algorithms for Maximum Lifetime Routing in Wireless Sensor Networks. In: Global Telecommunications Conference, pp. 2185–2193. IEEE Press, New York (2004)
10. Sankar, L.Z.: Maximum Lifetime Routing in Wireless Ad-hoc Networks. In: INFOCOM 2004, pp. 1089–1097. IEEE Press, New York (2004)

Estimation of Rock Mass Rating System with an Artificial Neural Network

Zhi Qiang Zhang^{1,2}, Qing Ming Wu¹, Qiang Zhang¹, and Zhi Chao Gong¹

¹ School of Power and Mechanical Engineering, Wuhan University, Wuhan 430072, China

² Hubei Provincial Key Laboratory of Fluid Machinery and Power Equipment Technology, Wuhan 430072, China

zhangzhiqiang78@263.net

Abstract. The geo-mechanical classification - rock mass rating (RMR) - is used for categorizing rock mass. Assessing RMR is an important factor for successful accomplishment of a tunneling project. In the rock mechanics and mining literatures, some empirical methods exist between rock mass and other rock properties, such as using characteristic of the rock, geological structure etc. However, those means have some limitations by special rock types. After analyzed the information to identify RMR, a new parameter as one of the input neurons was used to develop predictive relations. There are eight parameters as the input parameters are presented based on artificial neural networks (ANN). The situ-test data of the tunnel face were measured and the experimental results indicate the proposed method was effective.

Keywords: Rock mass rating (RMR), artificial neural network (ANN), tunnel boring machine (TBM), root mean square error (RMSE).

1 Introduction

Nowadays the trend in tunnel construction is toward larger and longer tunnels. Tunnel boring machine (TBM) is one of the important modern tunnel construction machines. However, in tunnel excavation by TBM, it is difficult to grasp the ground conditions ahead of and surrounding the tunnel face because the face cannot be observed during tunnel excavation [1]. In situ conditions, the environments of tunnel face are so complicated, hence fracture, faults and wet layers can be the limitation of the tunnel operation. With regard to the trend toward the use of TBM, the geological evaluation ahead of the tunnel face is an important issue to make effective use of it [2]. Most of rock mass rating systems assign numerical values to the different rock mass parameters that influence its behavior and thereafter combine these parametric values to give an overall rock mass rating (RMR) value [3]. Therefore, it is necessary to applying advancing geological prediction with RMR in order to reduce the risk of disasters.

As ANN model can cope with the complexity of intricate and ill-defined systems in a flexible and consistent way, in the last a couple of years an increase in their applications to solve various problems in the field of mechanics and mining geo-mechanics has been observed [4-8]. Chua and Goh [9] used Bayesian neural networks theory for

estimating wall deflections in deep excavations. A neural network system was developed by Javadi [10] for the estimation of air losses in compressed air tunnel.

2 Artificial Neural Network

The foundation of the artificial neural network (ANN) paradigm was laid in the 1950s. ANN model has non-linearity, high parallelism, robustness, fault and failure tolerance, learning, ability to handle imprecise and fuzzy information and so on [11]. Sarajedini [12] also indicates that ANN neurons are able to perform massively parallel computations for data processing and knowledge representation.

In this paper, a typical feedforward neural network (FNN) topology-- backpropagation network (BP) -- is introduced. It is comprised of the input layer, one or more hidden layers and the output layer. A topology of a simple FNN is presented in Fig. 1. Each layer includes a certain number of neurons that will transfer signals from one neuron to next.

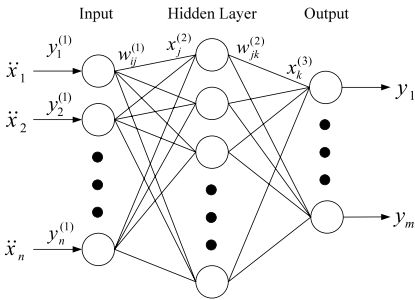


Fig. 1. A topology of neural network

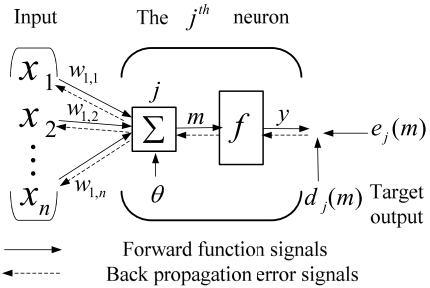


Fig. 2. The model of the j^{th} BP neuron and signal-flow

Fig. 1 also shows the process of the neuron feedforward.

$$\begin{cases} y_k^{(3)} = f \left(\sum_{j=0}^{n_1-1} w_{jk}^{(2)} x_j^{(2)} - \theta_k^{(2)} \right), & k = 0, 1, \dots, m-1 \\ x_j^{(2)} = f \left(\sum_{i=0}^{n_1-1} w_{ij}^{(1)} x_i^{(1)} - \theta_j^{(1)} \right), & j = 0, 1, \dots, n_1-1 \end{cases} \quad (1)$$

Clearly, equation (1) indicates that the neurons have responsibility for mapping n -dimension into m -dimension.

Fig. 2 describes a general model of one BP neuron, where x =input value; w =weight; Σ =summation; θ =bias; f =activation or transformation function and y =output value. One of the neurons' signal-flow is shown in equation (2, 3):

$$u_j = \sum_{i=1}^n w_i x_i - \theta_j \quad (2)$$

$$y_j = f(u_j) = \frac{1}{1 + e^{-x}} \quad (3)$$

where: y_j is the *sigmoid* function.

Based on the classic BP algorithm, the Levenberg-Marquardt algorithm is one of the most famous algorithms [7]. It depends on numerical optimization techniques to minimize and accelerate the required calculations, resulting in much faster training (Demuth and Beale, 1994). Fig. 2 also denoted the iterative process.

$$y_k(n) = y_{k-1}(n) - Y_{k-1}^{-1}(n) \cdot g_{k-1}(n) \quad (4)$$

Where, $Y_{k-1}(n)$ is the Hessian matrix of the error function at the current values of weights and biases and $g_{k-1}(n)$ is the gradient of the error function.

3 Parametric Establishment

It is very important to acquire the input parameters accuracy based on a sufficient number of training samples. For this purpose, Kavzoglu [13] and Gallagher [14] proposed that greater than 30 times the number of training samples as weights should be used. However, Staufner [15] suggested that the optimal number of training samples are almost 2/3 samples. While Lee [16] and Tong [17] proposed about 80% of data for training. Curry [18] recommended approximately 75% of data for training. In the present study, 120 component subset of the 150 component dataset (80% of database) were used in the training stage, and the remainder (30 components) was used in testing.

There are also some parameters (the initial weights, momentum coefficients (μ) and learning rate (η)) that can influence the BP network convergent accuracy. In the literatures, the initial weights are generally set small values. Different ranges were applied to set the initial weights: such as [-0.1; 0.1] by Staufner [15] and Rayburn [19]; [-0.25; 0.25] by Gallagher [14]; [-0.4; 0.4] by Weigend [20]; [-1; 1] by Looney [21]. Random small values are usually set as the initial weights which has a significant effect on both the convergence and final network architecture because too small range can result in small error gradients which may slow down the initial learning process. In this study, [-0.1; 0.1] was selected as the initial range.

The training rate of an ANN is also sensitive to the learning rate η and momentum coefficient μ . The larger of learning rate is selected, the quicker of the training rate, because large η value causes more changes to weights in the network. However the training phase can cause oscillations when η is selected too large. Lee [22] and Feng [23] set learning rate to 0.1, Tong [24] set it to 0.6 while Wang [25] endows it with 0.001. In order to solve learning rate sensitivity, μ is introduced. To a certain extent, the momentum coefficient μ has a stabilizing effect and makes curves smoothness. Curry [26] and Lee [22] set the momentum coefficient are 0.4 and 0.6 respectively; Feng [23] and Tong [24] suggested 0.1 and 0.9; and Wang [25] proposed 0.95. Therefore, in this study, the learning rate was selected as 0.1 and the momentum coefficient was set to 0.9.

What's more, the root mean square error (RMSE) is also important to ANN model, the learning rules are based on RMSE, the equation of RMSE is:

$$E = \frac{1}{2}(D - Y)^2 \quad (5)$$

Where: D and E are the output of expectation and RMSE respectively. The weight coefficient will be adapted by E and keep Y close to D . Obviously, when the signals flow forward to output layer and the results are compared with target output D . Fig. 2 also denoted the iterative process of one neuron.

It is necessary to have correct network architecture in order to reduce RMSE to minimum. Therefore the equation (5) is used as the transfer function in this study just as the most common transfer function implemented in the literatures [27-30]. In addition, the number of training neurons of the hidden layer is an important factor that will determine the training efficiency and optimization. Generally, excessive neurons of the hidden layer, which are also called over-fitting, can conduce near-zero error on predicting training data, or may lead to a longer training time and slower training speed and result in the process whereby the network learns the training data well but has no ability to meet results for test data. When training set size is too small, the network cannot learn effectively, this will lead to under-fitting and weak generalization. In a word, the appropriate number of hidden layer neurons and the minimal error of the test data are considered to be the optimal ANN architecture.

4 Design of the Optimal ANN

It is very important to verify the classification of the rock mass because the differentiation is the base of the TBM excavating and geological disaster prediction. Many scholars proposed some parameters of the rock mass that can decide excavation, for example: Benardos and Kaliampakos [7] proposed eight parameters as their ANN model for excavation by TBM. Suwansawat and Einstein [8] showed 13 input nodes for their ANN model. C.G. Chua and Goh [9] developed 35 input units in their paper. Palmstrom and Broch [31] considered 12 factors that influenced RMR. In this paper, several significant characteristics that influence the rock mass rating are adopted. These factors include:

- Rock mass quality (**Q**) represented by RMR classification;
- Characteristic of the rock mass (**RMC**);
- Hydrogeological conditions (**HC**) represented by the water surface relative to the tunnel;
- Geological structure (**GS**);
- Rock mass fracture degree (**RMFD**) as represented by rock quality designation (**RQD**);
- Weathering degree (**WD**) of the rock mass;
- Elastic longitudinal wave speed (V_n) of the rock mass.

Moreover, the value of *logsig* is salient between 0 and 1; the training sample should be processed normalizable data. The normalizable data of each parameter is presented in Table 2. The rock mass rating is made up of five grades [32] and normalizable data of each grade are also shown in Table 1.

Consequently, there are seven kinds of input values and five sorts of output values, namely $N_i=7$ and $N_o=5$. In order to obtain a good performance of the ANN, it is indispensable to have an optimal ANN model. The empirical calculated number of neuron of hidden layer is proposed by Table 3. Details on the implementation of this system are addressed in [4].

Table 1. Rock mass lithology description, classification and results

Rock mass lithology description	Rating	Results
Fault, chlorite schist, cataclasm and loose, weathering	I	(0.1, 0.9, 0.9, 0.9, 0.9)
Crash, dolerite, mid-weathering, ground water	II	(0.9, 0.1, 0.9, 0.9, 0.9)
Block crack, mid-weathering, ground water, mid-stiffness	III	(0.9, 0.9, 0.1, 0.9, 0.9)
Block, mid-weathering and fresh, without ground water, stiff	IV	(0.9, 0.9, 0.9, 0.1, 0.9)
Block, fresh, without ground water, stiff	V	(0.9, 0.9, 0.9, 0.9, 0.1)

Table 2. Normalizable data of the principal parameters

Q	types	horniness	middle	soft	----	----
	value	0.1	0.5	0.9	----	----
RMC	types	integrity	block	sandwich	chip	smash
	value	0.1	0.3	0.5	0.7	0.9
HC	types	dry	seep	drop	flow	stream
	value	0.1	0.3	0.5	0.7	0.9
GS	types	slight	less severity	severity	more	----
	value	0.1	0.4	0.7	0.9	----
RMFD	types	none	less growth	growth	more	----
	value	0.1	0.4	0.7	0.9	----
WD	types	none	slight	feeble	strong	full
	value	0.1	0.3	0.5	0.7	0.9
V_p (km/s)		V_p / V_p (max)			where, V _p (max)=5	
DR (s/20cm)	types	> 10	7~9	5~7	3~5	< 3
	value	0.1	0.3	0.5	0.7	0.9

As can be seen from Table 3, the number of hidden layer neuron, which calculated by empirical formula from 7 input neurons, varies between 6 and 21. The optimal ANN model will be established:

- Initial momentum coefficient $\mu=0.9$;
- learning rate $\eta=0.1$;
- Numbers of hidden layers: 1, 2;
- Numbers of hidden neurons in each hidden layer: 6, 10, 14, 21;
- The goal of the training: 0.01;
- The epochs of the training: 20000.

Table 3. The empirical calculated neuron of hidden layer(s) (N_i :number of input neuron, N_o :number of output neuron)

The empirical formula	input neurons=7	input neurons=8
$\leq 2 \times N_i + 1$	≤ 15	≤ 17
$3N_i$	21	24
$(N_i + N_o) / 2$	6	7
$\frac{2 + N_o \times N_i + 0.5N_o \times (N_o^2 + N_i) - 3}{N_i + N_o}$	10	10
$2N_i / 3$	5	6
$\sqrt{N_i \times N_o}$	6	6
$2N_i$	14	16

Fig. 3 shows the training RMSE and validation RMSE curve of 7 input neurons model, which including single-layer and double-layer in its hidden layer. Each ANN model is trained with the training set until it reaches pre-defined training goal. The parameters of validation set are consistent with corresponding model. The results are used for comparing with the desired outputs. If the outputs of the validation samples act in accordance with target data, the training will be finished. To evaluate the network architecture, each RMSE is used to be compared with other models in Fig. 3.

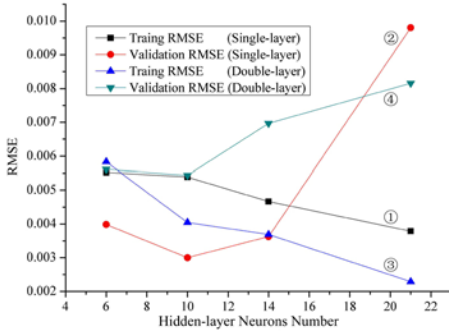


Fig. 3. The RMSE of 7 input neurons

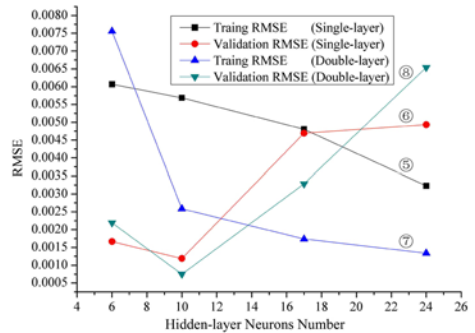


Fig. 4. The RMSE of 8 input neurons

The number of hidden neurons and layers are important variable that as a result is shown in Fig. 3 and Table 4 is definition of neural network models. Theoretically, the higher the number of hidden layers and neurons, the better the ANN fit the training data, just like curve ① and ③. But, more neurons may lead to “over-fitting”, which presented in the end of section 3, as shown in curve ② and ④. The larger number of the hidden layers and neurons failed to estimate rock mass rating. Therefore, model 2 is the optimal ANN model which has the highest prediction (88%) in these models.

5 A New Optional Parameter, Results and Discussion

In the previous section, the applicability of the 7 input neurons of neural network model for estimating rock mass classification was applied. The predictive model performed not well and confirmed that ANN can be unsuccessfully used for estimating RMR.

Determination of the advancing geological prediction by testing samples is almost impossible due to the presence of discontinuities. To overcome this difficulty, a lot of literatures [1-3, 33] have been proposed for predicting the rock mass rating. However, there are some limitations of these, the parameters, which defined by past, are not good parameters to predict RMR, namely parameters are insufficient. In view of the drilling machines are applied in locale, drilling rate (**DR**) is a factor not to be ignored too. So in this study, **DR** is set as another parameter, as can be seen from Table 2. Therefore, the number of optimal input parameters are eight and the hidden layer neurons number are also shown in Table 3. Based on the previous section, the optimal ANN model was founded. Only the hidden layer neurons have a change, all other parameters are unchangeable. That is to say, the numbers of hidden neurons in each hidden layer are changed into 6, 10, 17 and 24.

The results of the new neural network were given by Fig. 4. The training data of the new ANN model shows the double-layer neuron number are 10 by curve ⑦ and ⑧ is optimum (i.e., $RSME=0.002585$), the validation samples are tested by this model with lowest error (i.e., $RMSE=0.00075$). In addition, this model has higher efficiency and accuracy; in the meantime it consumes lower resource. Both the training and testing results of the network are plotted in Fig. 5.

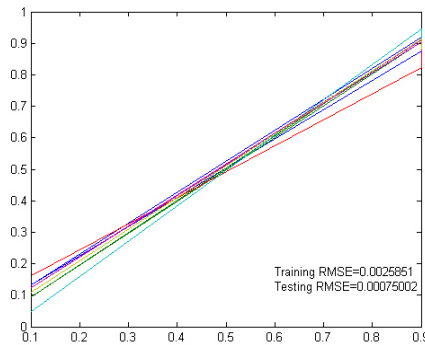


Fig. 5. Performances of the optimal ANN by model 14

As can be seen from Fig. 5, the new optional neural network model is able to establish a high correlation between the results of prediction and reality. The X-axis is normal results which are given by Table 1, while Y-axis is ANN predicted results. If the predicted results are very excellence, the curve will be a beeline with 45 degree. The linearity is better from this figure and the rate of the prediction is 92%. It is shown that the neural network is capable of capturing the main features of the relationship about the eight parameters and reflected the rule of the rock mass rating, it has higher reliability, so the rock mass rating of the tunnel face can be prediction accuracy.

Based on Table 4, seven other combinations of the hidden layers were examined when the successful ANN model 14 was established. In these models, the number of hidden layers and neurons was provided by Table 3. The architectures of these models were shown in Fig. 4. The results showed that, in all cases, the neural network was acted in accordance with pre-defined. For example, model 10 had 1 hidden layer and 10 hidden neurons. The neural network model saved time, had efficiency and accuracy. In addition, model 11 can also learn the prediction with high accuracy.

However there are some deficiencies why these models are not selected as the optimal ANN model. First of all, the RMSE is bigger than others' like model 11, 12, 16; Secondly, the prediction rating that maybe lower than others' such as model 9, 12, 13, 16, and another crucial point is over-fitting that will consume more resources and lower efficiency which is denoted by model 11, 12, 15, 16, and under-fitting which can make the network learn ineffectively just as model 9, 13. By comparing the results of these ANN models, model 14 is found to be optimal.

Table 4. Defined of neural network models

Model	Network architecture (7 input neurons)	Model	Network architecture (8 input neurons)
1	single-layer, 6 hidden neurons	9	single-layer, 6 hidden neurons
2	single-layer, 10 hidden neurons	10	single-layer, 10 hidden neurons
3	single-layer, 14 hidden neurons	11	single-layer, 17 hidden neurons
4	single-layer, 21 hidden neurons	12	single-layer, 24 hidden neurons
5	double-layer, 6 hidden neurons	13	double-layer, 6 hidden neurons
6	double-layer, 10 hidden neurons	14	double-layer, 10 hidden neurons
7	double-layer, 14 hidden neurons	15	double-layer, 17 hidden neurons
8	double-layer, 21 hidden neurons	16	double-layer, 24 hidden neurons

In a word, model 14 suggests that if a database exists for estimation of rock mass rating, a new ANN trained with this model can be applied to predict the RMR for a new project. It is obvious that the quality and results of prediction will improve with veracity of the RMR by tunnel excavation.

6 Conclusions

In order to fully grasp the classification of rock mass in the tunneling face, reduce the incidence of disaster, advancing geological prediction is necessary. With a view to the complexity and non-linear of the tunnel face, the RMR prediction takes full advantage of the artificial neural networks. In this study, a new parameter as input neuron method has been developed. In order to realize high performance in prediction of the rock mass rating, a four-layer BP network (8-10-10-5) is proposed, which is adopted as the optimal ANN model that is presented to update the neuron networks weights and momentum coefficient. The optimal model of ANN system demonstrates very satisfactory results in RMR prediction with locale data. The resulting remarks can be drawn hereinafter:

- A. This ANN system gives a fairly fast response for the RMR prediction.
- B. This ANN predictable scheme efficiently learns from situ-test data, the result of RMSE is lower than other models and achieves performance goal.
- C. A new parameter as the input neuron is proposed, it can improve the precision of the RMR prediction results in application.
- D. This optional ANN system can adapt rock mechanics and mining projects that are predicted the RMR before tunnel face. The results show the effectiveness of the presented control method.
- E. The open source code increases the optimal model's flexibility, allowing also the insertion of additional parameter to enhance the RMR prediction accuracy and efficiency.

Acknowledgments. This work is supported by the 10th Five-Year National Key Technological Equipment Plan of P.R. China (NO. ZZ02-03-03-02-02).

References

1. Shirasagi, S., Mito, Y., Aoki, K.: Evaluation of the Geological Condition Ahead of the Tunnel Face by Geostatistical Techniques Using TBM Driving Data. *Tunneling and Underground Space Technology* 18, 213–221 (2003)
2. Ashida, Y.: Seismic Imaging Ahead of a Tunnel Face with Three-component Geophones. *International Journal of Rock Mechanics & Mining Sciences* 38, 823–831 (2001)
3. El-Naqa, A.: Application of RMR and Q Geomechanical Classification Systems Along the Proposed Mujib Tunnel Route, Central Jordan. *Bull. Eng. Geol. Env.* 60, 257–269 (2001)
4. Sonmez, H., Gokceoglu, C., Nefeslioglu, H.A., Kayabasi, A.: Estimation of Rock Modulus: For Intact Rocks with an Artificial Neural Network and for Rock Masses with a New Empirical Equation. *International Journal of Rock Mechanics & Mining Sciences* 43, 224–235 (2006)
5. Karri, V.: Drilling Performance Prediction Using General Regression Neural Networks. In: Loganathara, R., Palm, G., Ali, M. (eds.) *IEA/AIE 2000. LNCS (LNAI)*, vol. 1821, pp. 67–73. Springer, Heidelberg (2000)
6. Lin, S.C., Ting, C.J.: Drill Wear Monitoring Using Neural Networks. *Int. J. Mach. Tools Manufact.* 36(4), 465–475 (1996)
7. Benardos, A.G., Kaliampakos, D.C.: Modeling TBM Performance with Artificial Neural Network. *Tunneling and Underground Space Technology* 19, 597–605 (2004)
8. Suwansawat, S., Einstein, H.H.: Artificial Neural Networks for Predicting the Maximum Surface Settlement Caused by EPB Shield Tunneling. *Tunneling and Underground Space Technology* 21, 133–150 (2006)
9. Chua, C.G., Goh, A.T.C.: Estimating Wall Deflections in Deep Excavations Using Bayesian Neural Networks. *Tunneling and Underground Space Tech.* 20, 400–409 (2005)
10. Javadi, A.A.: Estimation of Air Losses in Compressed Air Tunneling Using Neural Network. *Tunneling and Underground Space Technology* 21, 9–20 (2006)
11. Jain, A., Kumar, A.M.: Hybrid neural network models for hydrologic time series forecasting. *Applied Soft Computing* 7, 585–592 (2007)
12. Sarajedini, A., Hecht-Nielsen, R., Chau, P.M.: Conditional Probability Density Function Estimation with Sigmoidal Neural Networks. *IEEE Transactions on Neural Networks* 10(2), 231–238 (1999)
13. Kavzoglu, T., Mather, P.M.: Using Feature Selection Techniques to Produce Smaller Neural Networks with Better Generalisation Capabilities. In: *Geoscience and Remote Sensing Symposium*, vol. 7, pp. 3069–3071 (2000)

14. Gallagher, M., Downs, T.: Visualization of Learning in Multilayer Perceptron Networks Using Principal Component Analysis. *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics* 33(1), 28–34 (2003)
15. Stauffer, P.: *Spatial Analysis and GeoComputation*, pp. 183–207. Springer, Berlin (2006)
16. Lee, Y.C.: Application of Support Vector Machines to Corporate Credit Rating Prediction. *Expert Systems with Applications* 33, 67–74 (2007)
17. Tong, L.I., Chao, L.C.: Novel Yield Model for Integrated Circuits with Clustered Defects. *Expert Systems with Applications* 34(4), 2334–2341 (2007)
18. Curry, B., Morgan, P., Silver, M.: Neural Networks and Non-linear Statistical Methods: an Application to the Modeling of Price-quality Relationships. *Computers & Operations Research* 29, 951–969 (2002)
19. Rayburn, D.B., Klimasauskas, C.C.: The Use of Back Propagation Neural Networks to Identify Mediator-Specific Cardiovascular Waveforms. In: *International Joint Conference on Neural Networks*, vol. 2, pp. 105–110 (1990)
20. Weigend, A.S., Rumelhart, D.E., Huberman, B.A.: Generalization by Weight-Elimination applied to Currency Exchange Rate Prediction. In: *Seattle International Joint Conference on Neural Networks*, vol. 1, pp. 837–841 (1991)
21. Looney, G.G.: Advances in Feedforward Neural Networks: Demystifying Knowledge Acquiring Black Boxes. *IEEE Transactions on Knowledge and Data Engineering* 8(2), 211–226 (1996)
22. Lee, Y.C.: Application of support Vector Machines to Corporate Credit Rating Prediction. *Expert Systems with Applications* 33, 67–74 (2007)
23. Feng, X.Y., Wang, Q.Q., Zhang, J.: Studying Aromatic Compounds in Infrared Spectra Based on Support Vector Machine. *Vibrational Spectroscopy* 44(2), 243–247 (2007)
24. Tong, L.I., Chao, L.C.: Novel Yield Model for Integrated Circuits with Clustered Defects. *Expert Systems with Applications* 34(4), 2334–2341 (2007)
25. Wang, Y.S., Lee, C.M.: Sound-quality Prediction for Nonstationary Vehicle Interior Noise Based on Wavelet Pre-processing Neural Network Model. *Journal of Sound and Vibration* 299, 933–947 (2007)
26. Curry, B., Morgan, P., Silver, M.: Neural Networks and Non-linear Statistical Methods: an Application to the Modeling of Price-quality Relationships. *Computers & Operations Research* 29, 951–969 (2002)
27. Zhang, C.L., Mei, D.Q., Chen, Z.C.: Active Vibration Isolation of a Micro-manufacturing Platform Based on a Neural Network. *Journal of Materials Processing Technology* 129, 634–639 (2002)
28. Hecht-Nielsen, R.: Theory of the Backpropagation Neural Network. In: *International Joint Conference on Neural Networks*, vol. 1, pp. 593–650 (1989)
29. Sonmez, H., Gokceoglu, C., Nefeslioglu, H.A., Kayabasi, A.: Estimation of Rock Modulus: For Intact Rocks with an Artificial Neural Network and for Rock Masses with a New Empirical Equation. *International Journal of Rock Mechanics & Mining Sciences* 43, 224–235 (2006)
30. Karri, V.: Drilling Performance Prediction Using General Regression Neural Networks. In: Loganathara, R., Palm, G., Ali, M. (eds.) *IEA/AIE 2000. LNCS (LNAD)*, vol. 1821, pp. 67–73. Springer, Heidelberg (2000)
31. Palmstrom, A., Broch, E.: Use and Misuse of Rock Mass Classification Systems with Particular Reference to the Q-system. *Tunneling and Underground Space Technology* 21, 575–593 (2006)
32. Wang, L.K.: Application of Several Kinds of Advanced Forecast of Geology in Tunnel Construction. *Metal Mine* 305, 45–47 (2001) (in Chinese)
33. Banks, D.: Rock Mass Ratings (RMRs) Predicted from Slope Angles of Natural Rock Outcrops. *International Journal of Rock Mechanics & Mining Sciences* 42, 440–449 (2005)

Comparative Study on Three Voidage Measurement Methods for Two-Phase Flow

Youmin Guo¹ and Zhenrui Peng²

¹Institute of Mechatronic Technology, Lanzhou Jiaotong University,
Lanzhou 730070, China

²School of Mechatronic Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China
pengzr@mail.lzjtu.cn

Abstract. Voidage measurement of two-phase flow is of great importance to the industrial sector, in terms of safety, environmental protection and energy saving. Electrical capacitance tomography (ECT) is an effective technique for elucidating the distribution of dielectric materials inside closed pipes or vessels. Three currently proposed voidage measurement methods, which are based on Genetic Algorithm - Partial Least Square (GA-PLS), Ant System Algorithm (ASA), and Least Squares Support Vector Machine (LS-SVM) are reviewed respectively. All the data under three voidage measurement methods come from the ECT sensor. Then a unifying model is proposed to provide a universal voidage measurement framework. Finally, the experimental data are used to evaluate the three methods. The evaluation results are compared in terms of mean squared error, maximum absolute error, mean absolute error and measurement time. Future possible voidage measurement method based on ECT capacitance information is also discussed.

Keywords: Voidage measurement, Electrical Capacitance Tomography (ECT), Genetic Algorithm (GA), Ant System Algorithm (ASA), Least Squares Support Vector Machine (LS-SVM).

1 Introduction

Gas-liquid two-phase flow mostly exists in industries such as chemical, petroleum, and power industries, etc. Voidage is an important parameter of gas-liquid two-phase flow. The on-line voidage measurement has the advantages of safety, environmental protection and energy saving in industry. Owing to the phase interface and relative velocity of two phases, the flow characteristics of two-phase flow are far more complicated than that of single-phase flow. Hence, on-line voidage measurement has been a key problem in the two-phase flow research field. This problem has not been solved well till now [1-4].

Electrical Capacitance Tomography (ECT) technology, with features of simplicity, non-intrusion, low cost, and fast speed, has gained some achievements in the measurement of two-phase flow parameters including voidage and flow pattern. In the voidage measurement of two-phase flow, the commonly used method based on ECT is to reconstruct the cross-sectional image of two-phase flow and then obtain the voidage

value by calculating the grey level value of reconstructed image. Unfortunately, this method is very difficult to meet the requirements of high image reconstruction accuracy and good real-time performance simultaneously; hence, its practical application is also very limited [5-7].

12-electrode ECT system can obtain 66 measurement capacitances that reflect the phase fraction and distribution of two-phase flow [6]. It is possible for one to implement the voidage measurement without the complicated and time consuming image of reconstruction process if only one can find the correlations between the capacitances and the voidage. Based on this idea, some studies have been carried out. Recently, Wang [8] proposed a voidage measurement model based on the Genetic Algorithm (GA) and the Partial Least Square (PLS) methods, and this model is considered as a linear optimal capacitance combination for simplicity. Li et al [9] put forward another voidage measurement model which is also considered as a linear optimal capacitance combination, and is developed by using the Ant System Algorithm (ASA). Peng et al [10] introduced a voidage measurement method, in which Least Squares Support Vector Machine (LS-SVM) is used to establish the voidage measurement model.

In the above three methods, the voidage measurement model is established for each of three cross-section flow patterns respectively. The idea is to overcome the influence of the flow pattern on the voidage measurement. This is because, when the simple and fast linear back projection (LBP) algorithm is adopted to reconstruct the image for flow pattern identification, all the three methods will have the advantage of speed.

The aim of this study is to perform a comparative analysis of the above three voidage measurement methods to find the unifying framework for voidage measurement of gas-liquid two-phase flow. This perhaps, will be helpful for further research into new kinds of voidage measurement methods on the basis of capacitance information from ECT sensor.

2 Voidage Measurement Systems

Fig.1 shows the measurement system for the above mentioned three voidage measurement methods, which consists of three parts: 12-electrode ECT sensor, data acquisition unit and computer. The 12-elcotrode ECT sensor is composed of insulat

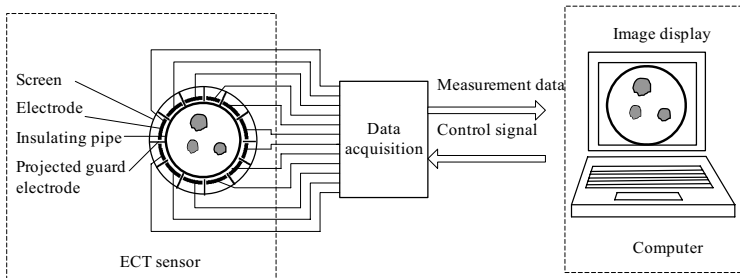


Fig. 1. Voidage measurement system

ing pipeline, projected guard electrodes, screen and 12 electrodes which are symmetrically mounted on the outside of an insulating pipe. The insulating pipeline of capacitance sensor is made of plexiglass. The pipeline is 50mm in inner diameter and is 4 mm in thickness. Each electrode is 60mm in length and 13 mm in width [11].

3 Three Voidage Measurement Methods

3.1 The Principle of Voidage Measurement

Each phase (component) of two-phase flow has its own permittivity (dielectric constant), which is different from that of the other phase. The change of phase fraction and distribution leads to the alteration of equivalent permittivity of two-phase flow and further results in the variation of capacitance values. By measuring the capacitance changes between all possible pairs of the electrodes of ECT sensor, the information of two-phase flow phase fraction and distribution can be obtained [7]. Then the voidage value $\tilde{\alpha}$ can be essentially regarded as a function of the measured capacitances.

$$\tilde{\alpha} = f(\mathbf{c}) = f(c_1, c_2, \dots, c_{66}). \tag{1}$$

where $\mathbf{c} = [c_1, c_2, \dots, c_{66}]^T$ is the normalized measured capacitance vector.

Due to complex relationship between the voidage and the capacitance values, there's no analytic expression of (1). In order to develop a voidage measurement model and estimate the voidage, Wang [8], Li et al [9], and Peng et al [10] proposed their methods for voidage measurement respectively.

3.2 The First Modeling Method of Voidage Measurement

Assume the voidage of two-phase flow can be computed by a linear combination of the measured capacitance values, and then the voidage measurement model can be expressed as below:

$$\tilde{\alpha} = \mathbf{w}\mathbf{c}. \tag{2}$$

where $\mathbf{c} = [c_1, c_2, \dots, c_{66}]^T$ is the normalized measured capacitance vector, and $\mathbf{w} = [w_1, w_2, \dots, w_{66}]$ is the coefficient vector, w_i is randomly taken as 1 or 0 at the beginning of the modeling (1 indicates the effective capacitance on the measurement voidage, and 0 indicates the ineffective capacitance.). The problem of exploring the optimal capacitance combination can be described by the following constrained optimization problem:

$$\begin{aligned} \min_{\mathbf{w} \in W} g(\mathbf{w}) &= \min_{\mathbf{w} \in W} |S_\alpha \alpha - \tilde{\alpha}| = \min_{\mathbf{w} \in W} |S_\alpha \alpha - \mathbf{w}\mathbf{c}| \\ \text{s.t. } w_i &\in \{0, 1\} \quad i = 1, 2, \dots, 66. \end{aligned} \tag{3}$$

where $g(\mathbf{w})$ is the objective function, α is the actual voidage, \mathbf{w} is the set of the coefficient vectors, S_α is the nonlinear scale transfer function of the voidage. The above optimization problem is solved by using Improved Genetic Algorithm (IGA).

Thus, the effective capacitances, which obviously contribute to the voidage measurement, and the ineffective capacitances, which have no significant contribution to the voidage measurement can be obtained. The Partial Least Square (PLS) method is further adopted to obtain the weight of the contribution of each effective capacitance. As a result, the voidage measurement model is finally developed as

$$\tilde{\alpha} = \mathbf{w}_f \mathbf{c} . \tag{4}$$

where $\mathbf{w}_f = [w_{f_1} \ w_{f_2} \ \dots \ w_{f_i} \ \dots \ w_{f_{66}}]$ is the regression vector, w_{f_i} is the weight of the contribution of the i th effective capacitance.

3.3 The Second Modeling Method of Voidage Measurement

In this method, the voidage of two-phase flow is also ideally assumed to be a linear combination of the measured capacitances. The voidage measurement model is in the same form as (2).

Obviously, different measured capacitances make different contributions to the voidage measurement. The Ant System Algorithm (ASA) is adopted to select the effective capacitances set from 66 measured capacitances obtained from the ECT sensor and also to determine the weight coefficient vector $\mathbf{w} = [w_1 \ w_2 \ \dots \ w_{66}]$. For each kind of flow patterns, the effective capacitance set and its contribution weight vector are determined by using the ASA.

3.4 The Third Modeling Method of Voidage Measurement

In this method, Least Squares Support Vector Machine (LS-SVM) is used to establish the voidage measurement models under different flow patterns. In each model, 66 capacitance values from ECT sensor are the inputs and the corresponding voidage value is the output. In the measurement process, the flow pattern of two-phase flow was identified first, and then the voidage is computed using the voidage model corresponding to the identified flow pattern.

Given l training data $(\mathbf{c}_1, \alpha_1), \dots, (\mathbf{c}_l, \alpha_l)$, the voidage regression model can be represented as the LS-SVM form

$$\tilde{\alpha}(\mathbf{c}) = \sum_{i=1}^l \beta_i K(\mathbf{c}, \mathbf{c}_i) + b . \tag{5}$$

where $\mathbf{c}_i \in R^{66}$ is the i th capacitance vector composed of 66 normalized capacitance values, $\alpha_i \in R$ is the actual voidage value in training data set, $\tilde{\alpha}$ is the measurement voidage (the estimated value) corresponding to \mathbf{c} , $\beta_i (i = 1, 2, \dots, l)$ is called the support vector coefficient (weight coefficient), b is the bias term, $K(\mathbf{c}, \mathbf{c}_i)$ is a kernel function satisfying Mercer's conditions.

The parameters $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_l)$ and b in (5) can be obtained by the following matrix equation

$$\begin{bmatrix} 0 & \mathbf{I}_v^T \\ \mathbf{I}_v & \boldsymbol{\Omega} + \frac{1}{\gamma} \mathbf{I} \end{bmatrix} \begin{bmatrix} b \\ \boldsymbol{\beta} \end{bmatrix} = \begin{bmatrix} 0 \\ \boldsymbol{\alpha} \end{bmatrix}. \quad (6)$$

where γ is the regulation parameter, $\mathbf{I}_v = [1; \dots; 1]$ is the column unit vector. Mercer's condition is applied within the matrix $\boldsymbol{\Omega}$

$$\Omega_{ij} = \beta_i \beta_j \varphi(\mathbf{c}_i)^T \varphi(\mathbf{c}_j) = \beta_i \beta_j K(\mathbf{c}_i, \mathbf{c}_j). \quad (7)$$

where $\varphi(\cdot)$ represents a high dimensional feature space, which is nonlinearly mapped from the input space.

3.5 A Unifying Framework for Three Methods above

Essentially, the three voidage measurement methods above can be formulated in a unifying framework. These three methods all take the capacitance information as the original signal. The analytical expressions can not be found between the voidage and the capacitance values because of the inherent complex characteristic of two-phase flow. The three methods mentioned above apply Genetic Algorithm (GA) and Partial Least Square (PLS), Ant System Algorithm (ASA), or Least Squares Support Vector Machine (LS-SVM) respectively to establish the voidage measurement models to perform voidage measurement. Essentially, the relationship can be considered as an optimization problem. Perhaps some proper heuristic, stochastic, or other kinds of optimization alternatives can be used for this task. From this point of view, other kinds of optimization methods could also be suitable for this task. For example, we are now using the Particle Swarm Optimization (PSO) to establish the voidage measurement model. The preliminary results are also satisfactory. Based on our research, we can depict a unifying modeling process of voidage measurement in Fig.2.

However, an important reservation is necessary. For the sake of simplicity and good real time performance, the first two methods have the same presumptions that there exists a linear combination between the measurement voidage (estimated voidage) value, and the measured capacitances $\tilde{\alpha} = \mathbf{w}\mathbf{c}$. The development of voidage measurement model is attributed to the problem of exploring the optimal combination. But there are differences between the first two methods. For the first method, IGA is selected to determine the optimal capacitance combination; PLS is to determine the weight of the contribution of each effective capacitance, while the latter adopts the ASA to determine the effective capacitance set \mathbf{c} and its weight vector \mathbf{w} . As for the third method, while establishing the voidage model, no linear relationship between the voidage values and capacitance values are assumed. From this respective, perhaps it is closer to the complex inherent of two-phase flow. The core idea of the third method is that of Support Vector Machine (SVM): map the training data nonlinearly into a higher-dimensional feature space and construct a separating hyper-plane with maximum margin there. This yields a nonlinear decision boundary in input space. By using one kernel function, it is possible to compute the separating hyper-plane without explicitly carrying out the map into the feature space.

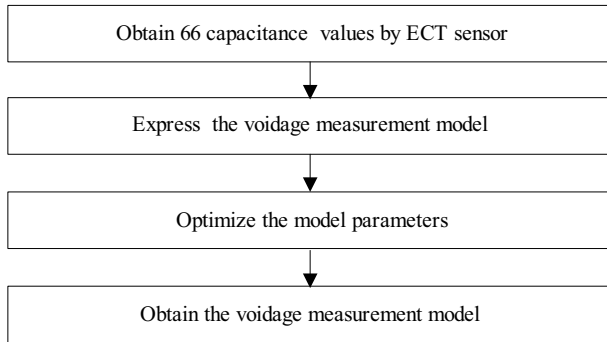


Fig. 2. The unifying modeling process of voidage measurement

4 Flow Pattern Identification

Flow pattern of two-phase flow is a spatial distribution of two phases. Because of the complexity of two-phase flow, it takes on several typical flow patterns such as bubble flow, plug flow, stratified flow, and annular flow. The flow pattern has great influence on the voidage measurement. It is very difficult for single voidage model to perform the voidage measurement under different flow patterns. Meanwhile, although the characteristics of two-phase flow are complicated, the voidage itself is a cross-section parameter, which is the area fraction of the cross section occupied by the gas phase and is determined by the two-dimensional distribution of two-phase flow. At a certain moment, the cross-section flow pattern only takes on one of the flow patterns (homogeneous flow, stratified flow and annular flow), which are shown in Fig. 3. For examples, bubble flow can be considered as homogeneous flow, the standard stratified flow and the wavy stratified flow can be treated as stratified flow, and slug flow can be regarded as the combination of stratified flow and homogeneous flow. Thus, for the measurement of voidage, the aim of the flow pattern identification is to identify the real-time flow pattern from homogeneous flow, stratified flow and annular flow [8, 12]. In this method, the flow pattern is identified by using fast linear back projection (LBP) image reconstruction and fuzzy pattern recognition technique. Flowchart for flow pattern identification is shown in Fig.4. The more detailed description can be found in [12].

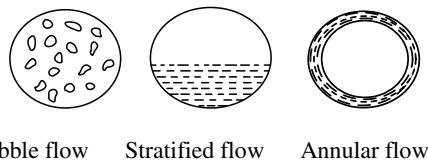


Fig. 3. Cross-section flow pattern in horizontal pipeline

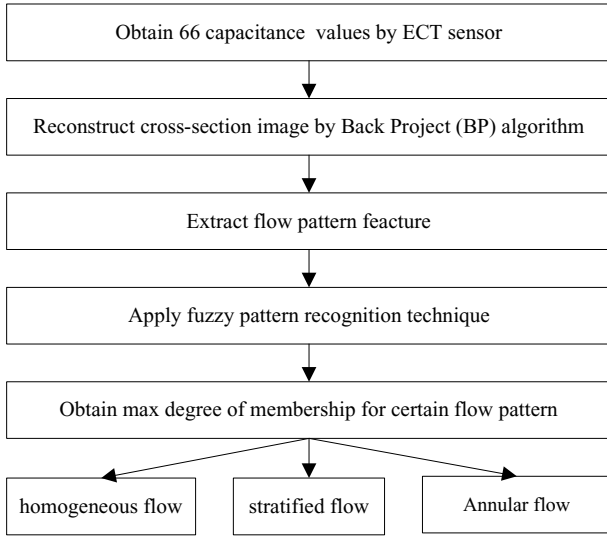


Fig. 4. Flowchart for flow pattern identification

5 On-Line Voidage Measurements of Three Methods

5.1 Unifying Voidage Measurement Process

The unifying flowchart for on-line voidage measurement process is shown in Fig. 5. Firstly, the capacitance measurement values are obtained by ECT sensor and then are normalized. Secondly, the crude cross-sectional image is reconstructed by using the fast BP algorithm. Thirdly, the flow pattern is identified by using fuzzy pattern recognition technique. Lastly, the normalized capacitance values are fed to a relevant established suitable model (established by any of the three methods) corresponding to the identified flow pattern to evaluate the voidage value.

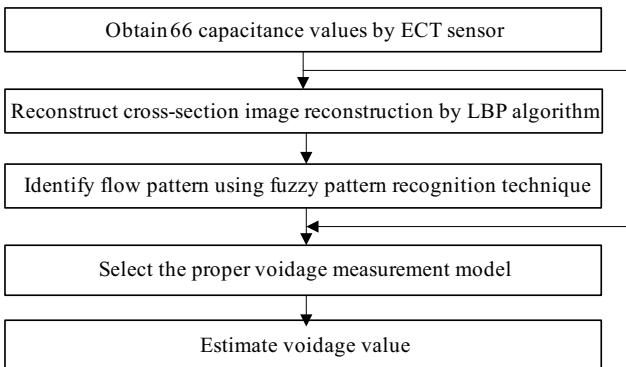


Fig. 5. Flowchart for the unifying voidage measurement process

5.2 Performance Criteria

The voidage measurement performance of each method is evaluated using the following statistical metrics, namely, the mean squared error (MSE), maximum absolute error (MaxE), mean absolute error (MeanE), and measurement time (t). The definitions of these criteria can be found in Table 1. MSE and MeanE are the measures of the deviation between the actual and measured voidage values. The smaller the values of MSE and MAE, the closer are measured voidage values to the actual voidage values. t is the measurement time. The smaller the value of t , the better real-time performance the voidage measurement has.

Table 1. Performance metrics and their calculations

Metrics	Calculations
MSE	$\frac{1}{k-1} \sum_{i=1}^k e_i^2$
MaxE	$\max\{ e_i \}, i = 1, 2, \dots, k$
MeanE	$\text{mean}\{ e_i \}, i = 1, 2, \dots, k$

k is the total number of data points. In our context, $k = 40$.

$$e_i = \tilde{\alpha} - \alpha$$

$\tilde{\alpha}$ and α represent the measured voidage value and actual voidage value respectively.

5.3 Experimental Results

Owing to the lack of effective dynamic voidage measurement method up to now, the static experiments are carried out to investigate the three voidage measurement methods above. In this experiment, we use gas, diesel oil and plexiglass tubes with different inner diameters to simulate the homogeneous flow, the stratified flow and the annular flow. The pipe is placed horizontally, and the diesel oil is partially filled to simulate the stratified flow. The plexiglass tubes with different inner diameters (simulating bubbles) are placed in the sensor, and then the diesel oil is filled into the test section to simulate the homogeneous flow. A thick plexiglass pipe is placed into the sensor, and the diesel oil is injected into the gap between the sensor and the plexiglass pipe to simulate the annular flow [8].

For each of the three methods above, numerous voidage measurement experiments of different flow patterns were carried out. The experimental results are listed in Table 2.

Table 2. Measurement results comparison of the three methods

Index	First method	Second method	Third method
MSE	0.0004371	0.0006612	0.0002538
MaxE	5.4%	5.6%	5.2%
MeanE	4.6%	4.8%	4.2%
t (s)	<0.09	<0.08	<0.08

The results show the maximum error of the voidage measurement is less than 6% with the three different methods and also demonstrate that the total voidage measurement time is less than 0.1s (The operating system was Windows XP Service Pack 2. The CPU of the computer was Intel Pentium (R) 4 CPU 2.60GHz and the memory was 512M). The accuracy and the real-time performance of the three voidage measurement methods can satisfy the field requirements.

Meanwhile, the third method has the least MSE and MaxE, which means the third method (LS-SVM method) has the best generalization ability. This perhaps attributes to the following two aspects. Firstly, LS-SVM model has no linear assumption, which may better accord with the complex characteristic of two-phase flow. Secondly, LS-SVM has good generalization ability itself by means of Structure Risk Minimum (SRM) principle, which seeks to minimize an upper bound of the generalization error consisting of both the training error and a confidence interval.

6 Discussions

This paper performs comparative study of three voidage measurement methods: GA-PLS, ASA, and LS-SVM based methods. The original data come from the ECT sensor. The measurement principle, the voidage measurement system, the flow pattern fuzzy recognition method, and the voidage measurement modeling methods proposed formerly were discussed first. Then the unifying voidage measurement modeling frame was established as well as the unifying voidage measurement process. Finally, the three voidage measurements were conducted and compared. It was found out that the third method (LS-SVM method) has the best generalization ability.

In short, we deduced a series of voidage measurement modeling methods for further research. Other modern optimization methods such as Particle Swarm Optimization (PSO) could also be suitable for the voidage measurement modeling. We also point out that it's possible to find out a new multi-model voidage measurement method, which can integrate these voidage measurement methods together to produce a more satisfactory voidage measurement results. All of these are very helpful for undertaking further research into new voidage measurement methods and further exploration of other parameters measurement methods of two-phase flow.

Acknowledgements

This paper is supported by the Support Program of Innovative Talents of Gansu Province (252003), the Natural Science Foundation of Gansu Province of China (No.3ZS062-B25-016), the Research Project of Gansu Education Department (20865) and "Qing Lan" Talent Engineering Funds by Lanzhou Jiaotong University. The authors would like to thank the reviewers for their helpful suggestions.

References

1. Li, H.Q.: Two-Phase Flow Parameter Measurement and Applications. Zhejiang University Press, Hangzhou (1991)
2. Hewitt, G.F.: Measurement of Two Phase Flow Parameters. Academic Press, London (1978)

3. Lin, Z.H.: Characteristics of Gas-Liquid Two-phase Flow in Pipelines and Their Engineering Applications. Xian Jiaotong University Press, Xian (1992)
4. Zhao, X., Jin, N.D., Li, W.B.: Soft Measurement Method of Phase Volume Fraction for Oil/ Water Two-Phase Flow. *Journal of Chemical Industry and Engineering (China)* 56, 1875–1879 (2005)
5. Marashdeh, Q., Fan, L.-S., Du, B., Warsito, W.: Electrical Capacitance Tomography - A Perspective. *Ind. Eng. Chem. Res.* 47, 3708–3719 (2008)
6. Li, H.Q., Huang, Z.Y.: Special Measurement Technology and Its Applications. Zhejiang University Press, Hangzhou (2000)
7. Huang, Z.Y., Wang, B.L., Li, H.Q.: Application of Electrical Capacitance Tomography to the Voidage Measurement of Two-Phase Flow. *IEEE Trans. Instrum. Meas.* 52(1), 7–12 (2003)
8. Wang, W.W.: Voidage Measurement of Gas-Oil Two-Phase Flow. *Chin. J. Chem. Eng.* 15, 339–344 (2007)
9. Li, Q.W., Huang, Z.Y., Wang, B.L., Li, H.L.: Void Fraction Measurement of Oil-Gas Two-Phase Flow Based on Ant System and Electrical Capacitance Tomography. *Journal of Chemical Industry and Engineering (China)* 58, 61–66 (2007)
10. Peng, Z.R., Wang, B.L., Huang, Z.Y., Li, H.Q.: Electrical Capacitance Tomography and LS-SVM Based Voidage Measurement. *Journal of Zhejiang University (Engineering Science)* 41, 877–880 (2007)
11. Wang, B.L., Ji, H.F., Huang, Z.Y., Li, H.Q.: A high-Speed Data Acquisition System for ECT Based on the Differential Sampling Method. *IEEE Sensors J.* 5, 308–311 (2005)
12. Xie, D.L., Huang, Z.Y., Ji, H.F., Li, H.Q.: An Online Flow Pattern Identification System for Gas-Oil Two-Phase Flow Using Electrical Capacitance Tomography. *IEEE Trans. Instrum. Meas.* 55, 1833–1838 (2006)

A New Approach to Improving ICA-Based Models for the Classification of Microarray Data

Kun-Hong Liu¹, Bo Li², Jun Zhang³, and Ji-Xiang Du⁴

¹ School of Software, Xiamen University, Xiamen 361005, Fujian, China

² School of Computer Science of Technology, Wuhan University of Science and Technology, 947 Heping Road, Wuhan 430081, Hubei, P.R. China

³ School of Electronic Science and Technology, Anhui University

⁴ Department of Computer Science and Technology, Huaqiao University, Quanzhou 362021, Fujian, P.R. China

Abstract. Inspired by the idea of ensemble feature selection, we design an ICA based ensemble learning system to fully utilize the difference among different IC sets. Firstly, some IC sets are generated by different ICA transformations. A multi-objective genetic algorithm (MOGA) is then designed to select different biologically significant IC subsets from these IC sets, which are applied to build base classifiers. In addition, a global-recording technique is designed to record the best IC subsets of each IC set discovered by the MOGA into a global-recording list. When MOGA stops, all individuals in the list are deployed to train base classifiers. The base classifiers generated by these schemes are fused by the majority vote rule. Three microarray datasets are used to test the ensemble systems, and the corresponding results demonstrate that two ensemble schemes can improve the performance of the ICA based classification model.

1 Introduction

Recent years, independent component analysis (ICA) transformation has been applied to the analysis of microarray data with great success, and there have been many algorithms and methodologies based on ICA proposed to analyze microarray data [1-7]. In these papers, the authors mainly paid attention to the biological interpretation of ICA results, but the discussions on how to select proper independent components (ICs) for different prediction systems are weak or completely ignored. However, it was found that the dominant ICs are related to particular biological or experimental effects, and the component weights are either tumor cluster or chromosomal aberration specific [1, 6]. So by using proper IC subsets, the performance of ICA based prediction system will be further improved.

But up to now, no universal rule for IC selection is available. The reasons lie in some aspects: firstly, neither the energies nor the biological significance of different ICs can be determined immediately, so the simple principle for principal component (PC) selection in principal component analysis (PCA) transformation can't be applied to IC selection. Secondly, different ICA algorithms are designed based on different estimate rules and objective functions, so they will generate different IC sets even for

a same source data. Thirdly, the results obtained from an ICA algorithm are not “ordered”. In short, it is impossible to set up a simple and universal rule to guide the IC selection for all classification systems. But similar to the feature selection problem, an efficient selection algorithm can be used to select a proper IC subset from an IC set for prediction efficiently. And in [8, 9], the sequential floating forward selection (SFFS) and genetic algorithm (GA) were used to deal with the IC selection problem for different ICA based models successfully. With these methods, an optimal IC subset will be selected to build an accurate classifier.

Unlike feature selection methods, ensemble feature selection (EFS) is a more efficient method to construct multiple classifier system (MCS) [10], which is implemented by training base classifiers with different feature subsets. Evolutionary based methods build base classifiers using optimal feature subsets selected by evolutionary approaches and can usually achieve stable and accurate results [11-13]. Inspired by the EFS, it is obvious that when building base classifiers with different IC subsets extracted from different IC sets, high diversity among the base classifiers will be achieved. And it has been proved that a MCS is more robust than an excellent single classifier in many fields [10]. So the ensemble learning scheme for ICA is promising, and is named as ensemble independent component selection (EICS) here.

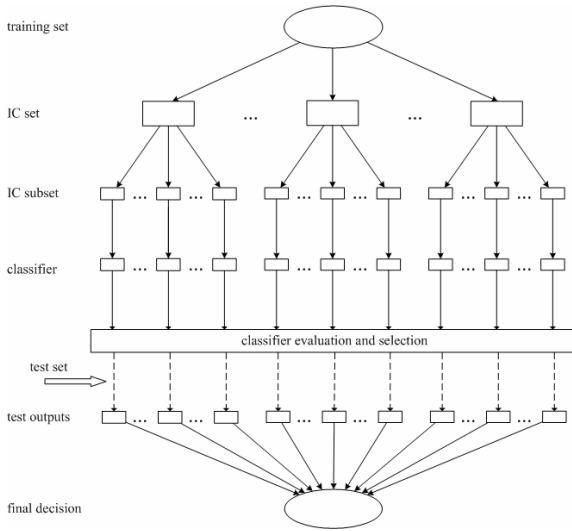


Fig. 1. The process of ensemble independent component selection

Due to thousands of gene data in a microarray dataset, the ICA algorithms require searching the maxima of a target function in a high dimensional space. As a result, most algorithms may get stuck in local maxima in the process of searching the maxima, and ICA is not always reproducible when used to analyze gene expression data even using a same ICA algorithm [1]. Moreover, the results may be sensitive to the initializations, which is still an unsolved problem up to now. Instead of solving this problem, we design an ensemble IC selection system by utilizing the difference among the generated IC sets. The framework of our EICS is illustrated by Figure 1. In

this system, some different IC sets are generated by different ICA transformations with random initializations firstly, and then an algorithm is deployed to search some biologically significant IC subsets from each IC set. The selected IC subsets are used to build classifiers, which are then selectively combined to construct a MCS. In this scheme, the diversity among the IC sets is no longer a trouble problem to deal with. On the contrary, it benefits the EICS scheme because it provides an easy method to maintain the diversity among base classifiers, which is important to the generalization capability of an ensemble system. Because of the stability of PCA algorithm, this ensemble method can't be applied to the PCA based prediction methods.

When implementing the EICS, a multi-objective genetic algorithm (MOGA) based scheme is designed with the goals to minimize error rate of classifiers and maximize the covering of IC sets at the same time. And a global-recording technique is designed to record the best two IC subsets to a global-recording list for each IC set discovered in evolution. After the MOGA steps, two fusion schemes are used to combine the classifiers: combining all individuals and combining the individuals above average accuracy in the list. When testing them on three microarray datasets, we find that these schemes are efficient and effective.

This paper is organized as follows. Section 2 presents the ICA based microarray dataset classification model. The framework of the MOGA based EICS is described in Section 3. In Section 4, experimental results are shown along with corresponding discussions. Then Section 5 concludes this paper.

2 ICA Based Prediction Model

Assume an $n \times p$ data matrix X , whose rows $r_i (i=1, \dots, n)$ correspond to observational variables and whose columns $c_j (j=1, \dots, p)$ are the individuals of the corresponding variables. Then the ICA model of X is:

$$X=AS \tag{1}$$

Without loss of generality, A is an $n \times n$ matrix, and S is an $n \times p$ source matrix whose rows are as statistically independent as possible. Those variables in the rows of S are ICs, and the statistical independence between variables is quantified by mutual information $I=\sum H(S_k)-H(S)$, where $H(S_k)$ is the marginal entropy of the variable S_k , and $H(S)$ is the joint entropy. And the ICs are estimated by:

$$U=S=A^{-1}X=WX \tag{2}$$

Let matrix X denote the gene expression data, then it is described as a linear mixture of statistically independent basis snapshots (eigenassay) S combined by an unknown mixing matrix A . In this approach, ICA is used to find a weight matrix W such that the rows of U are as statistically independent as possible. The independent eigenassays estimated by the rows of U are then used to represent the snapshots. The representation of the snapshots consists of their corresponding coordinates with respect to the eigenassays defined by the rows of U , i.e.,

$$r_j = a_{j1}u_1 + a_{j2}u_2 + \dots + a_{jn}u_n \tag{3}$$

The original training data sets X_{in} and test data sets X_{tt} are transposed so that they will be applied to evaluate the ICs with the following formulae:

$$U=W_{in}X_{in}=A_{in}^{-1}X_{in} \tag{4}$$

$$X_{in}=A_{in}U \tag{5}$$

The rows of A_{in} contain the coefficients of the linear combination of statistical sources that comprise X_{in} . Then the representation of the test set X_{it} is calculated as:

$$A_{it}=X_{it}U^{-1} \tag{6}$$

And after selecting some special ICs, formulas (1-6) are still applicable by adjusting A_{in} as $n \times m$, S as $m \times p$ and A_{it} as $k \times m$ if there are m ICs selected. Then the ICA model is constructed based on the selected ICs. In detail, after an ICA transformation, an IC subset is selected from the IC set to construct an IC subspace. Then a classifier is trained and then used to classify new samples in this IC subspace.

In this study, we first employ FastICA [15] on gene expression data to generate 50 different IC sets in our experiments.

3 The Design of MOGA

In the design of EICS, different IC subsets are selected from each IC set, and then are used to build base classifiers. It is obvious that an efficient selection method is vital to an efficient ensemble system. The rise of GA is inspired by the mechanism of evolution in nature. Compared with the sequential search algorithms, such as SFFS, GA has great advantages over them in the search of feature subsets because it always evaluates a subset as a whole, which is presented as a chromosome. Furthermore, many GA based ensemble systems have been proposed to tackle the classification problems with great success [11-13, 16-18]. So GA is deployed to implement the ensemble scheme here.

The framework of the GA based ensemble scheme is outlined as follows. Binary coding scheme is applied. Each chromosome represents an IC subset, and is comprised with two parts: the index of IC set and the mask for IC subset selection, as illustrated in Figure 2. If there are N IC sets, the length of the first part is calculated by $\lceil \log_2(N) \rceil$, which means the minimum integer larger than $\log_2(N)$. In the decoding process, these bits are converted to an integer, which indicates the index of the selected IC set. The length of the second part is equal to the number of ICs, which may vary for different microarray datasets. At this part, each gene is valued as 1/0 to represent whether a corresponding IC is/isn't selected. In this scheme, each chromosome represents an IC subset.

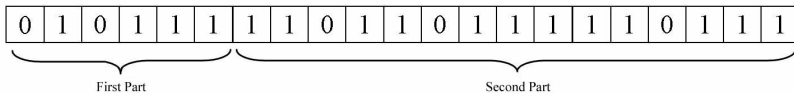


Fig. 2. The chromosome design scheme of the GA

The selection operator is roulette, which allows chromosomes with low fitness value to get a chance to enter the next generation. Double point recombination operator is adopted to exchange a randomly selected part of individuals in pairs. The simple inversion mutation is employed as the mutation operator, and it randomly selects two

points in a parent and produces offspring by reversing the genes between the two points. These operators guarantee the diversity among the population, which is important to an ensemble system.

It is obvious that if the base classifiers are built based on IC subsets selected from different IC sets, high diversity among these classifiers would be easily achieved. So the GA should evolve towards two desired goals: minimizing the error rate and maximizing the covering of IC sets. The fitness design scheme used in [16] is based on the idea of sharing method. Inspired by this scheme, we apply a multi-objective GA (MOGA) here. The first objective is to generate accurate individuals, and the second is to encourage the covering of the IC sets. In this way, both the covering and the accuracy are achieved at the same time.

Bootstrap 632+ is a widely deployed method for the estimation of generalization error [19]. Although it is time-consuming, the final results are close to be unbiased for small sample size problem. It is deployed to evaluate the performance of the base classifiers built by selected IC subsets in each generation. And the first optimization goal for the MOGA is to minimize the bootstrap .632+ error rates.

The IC subsets represented by the individuals in a population is called as a group here if they belong to a same IC set. In order to encourage the covering, we use the method similar to the sharing scheme [16]. That is, if there are n IC subsets in a group, the fitness values of these IC subsets are multiplied by $1/n$. Assuming two individuals, x_i and x_j , belong to the same group. The Hamming distance is used to evaluate the difference between them. The second fitness function for an individual is the sum of its Hamming distance in its group. When there is only a single subset in a group, the Hamming distance can't be calculated, and we simply assign a very large value, for example, 1000, to it. So this fitness value is calculated by (7). According to this formula, if a group is of small scale, all individuals in this group will have high fitness value at the second goal. This optimization goal is used to encourage the covering of IC sets by keeping the size of group small in each population.

$$F_i = \begin{cases} \sum_{j=1, j \neq i}^n \text{xor}(x_i, x_j) / n, & \text{if } n > 1 \\ 1000, & \text{if } n = 1 \end{cases} \quad (7)$$

Due to the encouragement of diversity, it is of great probability that some accurate individuals will be replaced in evolution if they are in the same group. If an ensemble is constructed with accurate individuals coming from all different IC sets, the difference among the IC sets will be fully utilized. So a global-recording technique is designed in our MOGA to record the best IC subsets of each IC set globally. In detail, a global-recording list is used to record the best two subsets of each IC set obtained during evolution. By combining the individuals in the list, an efficient ensemble system is built, which is denoted by EICS-1. In the evolution, the global-recording list may be updated in each generation by replacing the IC subsets in the list with the IC subsets that achieve lower error rate in the current generation. It should be noted that only the subsets belonging to the same IC set are compared, and only the best two subsets for each IC set are kept in the list.

The framework of this MOGA is based on NSGA-II [20], which implements a diversity-preserving mechanism. The chromosomes in a population are first sorted based on the nondominated sorting, and are assigned to different ranks according to the individuals dominated by them. Then they are assigned to a crowding distance

according to the difference among objective values. The selection is performed using the crowded tournament selection. In this process, the chromosomes with lower rank and larger crowding distance are chosen.

4 Experimental Results and Discussions

Three publicly available microarray datasets are deployed in our experiments: the hepatocellular carcinoma dataset [21], the prostate cancer dataset [22] and the breast cancer dataset [23]. In these datasets, all samples have already been assigned to the training set or test set, and Table 1 summarizes these datasets. Preprocessing of the datasets is done exactly as [24]: transforming the raw data to natural logarithmic values, and then standardizing each sample to zero mean and unit variance. Besides the original division, each dataset are reshuffled with 9 randomizations. And for all datasets, each randomized training and test set contains the same amount of samples of each class compared to the original training and test set. In all our experiments, the classifiers are built using the training samples, and the classification accuracies are estimated using the independent test set.

Table 1. The summary of the datasets

Data sets	training set	test set	Levels	microarray technology
Hepatocellular	33	27	7129	oligonucleotide
Prostate	102	34	12600	oligonucleotide
breast	78	19	24188	cDNA

For ICA based microarray dataset analysis, it is still an unsolved problem that how many ICs should be generated after an ICA transformation. One widely used method is to set the number of ICs to the number of samples due to the small training sample size of microarray datasets, as done in [5, 9]. This method is applied to the hepatocellular dataset. But for the prostate and breast cancer datasets whose training sample sizes are 78 and 102, respectively, it takes quite a long time to transform the datasets with so many ICs. In addition, the larger number of ICs will make it more difficult to search global optimal results. So we simply set the number of ICs to 50 for these two datasets to simplify our discussion.

According to the chromosome design scheme, six bits are used at the first part of chromosomes because there are fifty different IC sets in our experiments. As the population size is 100, in the first generation, the integers represented by the first part of chromosomes take value in the range of [1, 50] in sequence for the first and the second fifty individuals. In this way, all the IC sets appear twice in the first generation, so that they have an equal chance to compete with others at the initial stage. The second part of each chromosome representing the IC mask is randomly initialized. In experiments, for original division or each random initialization on each dataset, the MOGA runs five independent times. So in all, the results obtained by the MOGA are based on 50 runs for each dataset.

The bootstrap sample size is set to 100 in all experiments. Despite of the relatively small bootstrap size, bootstrap .632+ is quite time consuming when used in MOGA

for the fitness value evaluation. Based on this consideration, although some accurate classifiers, such as neural network and support vector machine, can also be employed to deal with the classification task, only the nearest neighbor classifier (1-NN), a relatively weak learner, is deployed in all our experiments owing to its small computational cost. And it should be noted that this method is independent of base classifier. When more accurate prediction systems are deployed as base classifiers, better performance will be achieved using our method.

Table 2. The prediction results on test sets based on fifty independent runs for each dataset. In the table, Aver_1 represents the average prediction results in the final generation, and Aver_2 represents the average prediction results in the global-recording list.

method		Hepatocellular	Prostate	Breast	
LS-SVM linear kernel		68.43±4.52	84.31±13.66	67.92±8.58	
LS-SVM RBF kernel		68.61±6.32	88.10±4.93	68.42±7.62	
LS-SVM linear kernel (no regularization)		49.56±12.60	48.18±10.25	57.14±9.08	
PCA + FDA (unsupervised PC selection)		68.25±7.37	83.89±13.63	57.39±15.57	
PCA + FDA (supervised PC selection)		66.67±9.96	82.49±13.35	66.92±9.90	
kPCA lin + FDA (unsupervised PC selection)		68.25±7.37	85.01±9.07	60.90±14.49	
kPCA lin + FDA (supervised PC selection)		66.67±9.96	82.49±13.35	65.41±7.54	
kPCA RBF + FDA (unsupervised PC selection)		61.20±12.91	85.01±11.00	51.38±15.91	
kPCA RBF + FDA (supervised PC selection)		69.49±3.94	28.71±10.02	36.84±0.00	
ICA+1-NN		66.68±7.15	92.06±8.26	65.63±7.33	
MOGA based ensemble IC selection	30 generations	EICS-1	63.11±3.60	99.71±0.93	68.42±4.30
		EICS-2	61.52±2.34	99.71±0.93	78.42±4.61
		Aver_1	58.78±8.53	91.50±5.00	64.89±10.04
		Aver_2	58.67±9.43	88.41±8.42	59.53±11.39
	50 generations	EICS-1	63.74±3.12	99.71±0.93	70.53±6.18
		EICS-2	62.89±2.73	99.71±0.93	78.42±6.30
		Aver_1	57.89±8.12	91.24±5.35	65.32±9.93
		Aver_2	58.44±9.73	88.62±8.15	60.11±11.23
	100 generations	EICS-1	64.89±3.68	100±0.00	72.11±6.10
		EICS-2	64.19±4.35	99.82±1.52	82.11±5.08
		Aver_1	58.70±6.08	90.82±4.83	65.89±8.93
		Aver_2	59.05±9.10	88.94±7.90	61.11±11.39
	150 generations	EICS-1	67.63±4.08	100±0.00	74.74±6.47
		EICS-2	69.75±5.05	99.24±1.52	79.47±5.79
		Aver_1	59.81±6.15	90.41±5.08	68.68±8.96
		Aver_2	59.56±8.74	89.76±7.98	61.95±11.09
	200 generations	EICS-1	68.52±4.92	99.41±1.24	75.79±4.44
		EICS-2	70.15±6.50	99.65±2.32	82.11±4.44
		Aver_1	59.11±5.62	90.26±5.55	68.79±9.41
		Aver_2	61.44±8.16	90.06±7.27	62.63±11.13

The results shown in Table 2 are the mean and standard deviation of the results based on each original dataset and 9 randomizations. For comparison, we list the results using 9 different methods in Table 2: PCA and kernel PCA with FDA, LS-SVM [24]. These corresponding results were all obtained based on a single classifier. We list the results obtained by an ICA based 1-NN. From Table 2, it is found that when only comparing the results obtained by a single classifier, ICA based 1-NN is not superior to the other nine methods, and no method can lead to obvious advantage in classification. So it is hard to choose a best method for the classification of all these three microarray datasets, which reveals the limit of the single classifier system.

It is found that by setting the covering of IC sets as one optimization goal, there are usually about twenty different IC sets in each generation. And in evolutionary process, there are less than twenty IC sets appearing with high frequency usually because of the random search mechanism of the MOGA. As analyzed before, once a group contains some more individuals picked by random style, the individuals in this group can get more opportunities to be optimized on the first goal despite of getting relatively low scores on the second goal. Then only the IC subsets obtained from the IC sets with high appearing frequency would be fully optimized.

From Table 2, it is interesting to find that although the average test results of base classifiers for each ensemble system are slightly worse than those obtained by a single classifier using a whole IC set for classification, the ensemble based classification results are much better than all others in most cases. The success of the ensemble systems lies in the diversity among the base classifiers. As different base classifiers produce different outputs by projecting samples into different IC subspaces, the samples that can't be correctly classified in an IC subspace may be recognized in other subspaces. Then even when some base classifiers make wrong decision on a sample, other classifiers can still have a chance to correct this result. So the ensemble system can produce a correct output. In this way, the final results are better than both the average results of base classifiers and the results of using a single classifier.

With the global-recording technique, the best subset of each IC set found by far will be recorded. Due to the computational cost, the MOGA only runs a relatively small number of generations and explores a very small part of the search space, so it is impossible for the MOGA to fully investigate all the IC sets. As proved in [18], by pruning some classifiers in an ensemble system, a compact and still accurate (or even more accurate) ensemble system will be built. And the idea of "overproduce and choose" has been proved to be successful in the design of ensemble systems. In [13, 18], after generating a set of base classifiers, the authors applied GA to choose the best team of classifiers to build ensemble systems. However, it is obvious that it requires much more time to apply this method for selective combination. An alternative method is to apply the individuals with above average accuracy in the global-recording list to construct a compact ensemble system, and the diversity is maintained by the global-recording list. With this method, about fifty individuals will be selected, which are originated from about twenty IC sets. As the number of further optimized IC sets is close to twenty in the evolutionary process, by keeping the individuals above average classification accuracy, all (or at least most of all) of the top individuals are included. This ensemble scheme is denoted by EICS-2, which could achieve good performance with smaller ensemble size.

The results of EICS-1 and EICS-2 are stable and usually keep increasing in evolution. From Table 2, it is obvious that EICS-1 can't always guarantee the best performance except for the prostate dataset. When only running a small number of generations, for example, 30 generations, the corresponding results are not always good enough. For EICS-1, the performance is mainly affected by the accuracies of the base classifiers. As the base classifiers in the global-recording list are more and more accurate during the evolution, EICS-1 can achieve better and better performance. And it also holds for EICS-2. When running some more generations, the results of EICS-1 and EICS-2 will be much better, as validated by the results shown in Table 2. But

since the results obtained in 200 generations are usually satisfying, we do not further evolve the MOGA.

It is found that there are about 40 base classifiers in EICS-2 in all experimnts. In comparison, there are always 100 base classifiers in EICS-1. Then usually EICS-2 is more efficient than EICS-1 by achieving close or higher classification accuracy with fewer base classifiers, which proves that the simple pruning method works very well. So EICS-2 is the best choice for the microarray dataset classification.

5 Conclusions

For microarray datasets, different IC sets would be generated after different initialized ICA transformations. So it is an unsolved problem that how to obtain a set of stable ICs for microarray dataset analysis. But in this study, the diversity among different IC sets is utilized to design a MOGA based ensemble system. The MOGA evolves towards two goals: minimizing the error rate and maximizing the covering of IC sets. A global-recording technique is designed, which records the best two results for each IC set in a global-recording list. Two ensemble schemes are implemented to fuse the base classifiers generated by the MOGA: combining all individuals in the final generation and only the individuals above average accuracy in the global-recording list. Compared with the results obtained by other methods on three microarray datasets, it is found that our ensemble schemes are effective and efficient in classifying normal and tumor samples from the three human tissues. In conclusion, it is obvious that the study of ensemble IC selection system is just at its beginning stage. In future works, we would apply ensemble systems based on different ICA classification models, and try to solve other practical problems.

Acknowledgments. This work was supported by the grants of the National Science Foundation of China (No.60805021 and 60772130), the China Postdoctoral Science Foundation (No.20060390180 and 200801231), and the grants of Natural Science Foundation of Fujian Province of China (No.A0740001 and A0810010).

References

1. Liebermeister, W.: Linear modes of gene expression determined by independent component analysis. *Bioinformatics* 18, 51–60 (2002)
2. Zhang, X.W., Yap, Y.L., Wei, D., Chen, F., Danchin, A.: Molecular Diagnosis of Human Cancer Type by Gene Expression Profiles and Independent Component Analysis. *European Journal of Human Genetics* 13(12), 1303–1311 (2005)
3. Zheng, C.H., Chen, Y., Li, X.X., Li, Y.X., Zhu, Y.P.: Tumor classification based on independent component analysis. *International Journal of Pattern Recognition and Artificial Intelligence* 20(2), 297–310 (2006)
4. Lee, S.I., Batzoglou, S.: Application of independent component analysis to microarrays. *Genome Biol.* 4(R76) (2003)
5. Huang, D.S., Zheng, C.H.: Independent component analysis-based penalized discriminant method for tumor classification using gene expression data. *Bioinformatics* 22, 1855–1862 (2006)

6. Frigyesi, A., Veerla, S., Lindgren, D., Hoglund, M.: Independent component analysis reveals new and biologically significant structures in microarray data. *BMC Bioinformatics* 7, 290 (2006)
7. Chiappetta, P., Roubaud, M.C., Torresani, B.: Blind source separation and the analysis of microarray data. *Journal of Computational Biology* 11, 1090–1109 (2004)
8. Zheng, C.H., Huang, D.S., Shang, L.: Feature selection in independent component subspace for microarray data classification. *Neurocomputing* 69(16-18), 2407–2410 (2006)
9. Liu, K.H., Huang, D.S., Zhang, J.: Improving the Performance of ICA Based Microarray Data Prediction Model with Genetic Algorithm. *IEEE CEC 2007*, 606–611 (2007)
10. Kuncheva, L.I.: *Combining pattern classifiers: methods and algorithms*. Wiley, Chichester (2004)
11. Opitz, D.: Feature selection for ensembles. In: *Proceedings of 16th National Conference on Artificial Intelligence (AAAI)*, pp. 379–384 (1999)
12. Kuncheva, L.I., Jain, L.C.: Designing classifier fusion systems by genetic algorithms. *IEEE Transactions on Evolutionary Computation* 4(4), 327–336 (2000)
13. Oliveira, L.S., Morita, M., Sabourin, R.: Feature selection for ensembles using the multi-objective optimization approach. In: *Studies in computational intelligence*, vol. 16, pp. 49–74. Springer, Heidelberg (2006)
14. Ho, T.K.: The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(8), 832–844 (1998)
15. Hyvärinen, A.: Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Netw.* 10, 626–634 (1999)
16. Liu, Y., Yao, X., Higuchi, T.: Evolutionary Ensembles with Negative Correlation Learning. *IEEE Transactions on Evolutionary Computation* 4(4), 381 (2000)
17. Wang, X., Wang, H.: Classification by evolutionary ensembles. *Pattern Recognition* 39(4), 595–607 (2006)
18. Zhou, Z.H., Wu, J., Tang, W.: Ensembling neural networks: Many could be better than all. *Artificial Intelligence* 137(1-2), 239–263 (2002)
19. Efron, B., Tibshirani, R.J.: Improvements on cross-validation: the 632+ bootstrap method. *J. Am. Stat. Assoc.* 92, 548–560 (1997)
20. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation* 6(2), 182–197 (2002)
21. Iizuka, N., Oka, M., Yamada-Okabe, H., Nishida, M., Maeda, Y., Mori, N., Takao, T., Tamesa, T., Tangoku, A., Tabuchi, H.: Oligonucleotide microarray for prediction of early intrahepatic recurrence of hepatocellular carcinoma after curative resection. *The Lancet* 361, 923–929 (2003)
22. Singh, D., Febbo, P.G., Ross, K., Jackson, D.G., Manola, J., Ladd, C., Tamayo, P., Renshaw, A.A., D'Amico, A.V., Richie, J.P., Lander, E.S., Loda, M., Kantoff, P.W., Golub, T.R., Sellers, W.R.: Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* 1(2), 203–209 (2002)
23. van't Veer, L.J., Dai, H.Y., van de Vijver, M.J., He, Y.D.D., Hart, A.A.M., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., Schreiber, G.J., Kerkhoven, R.M., Roberts, C., Linsley, P.S., Bernards, R., Friend, S.H.: Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415(6871), 530–536 (2002)
24. Pochet, N., De Smet, F., Suykens, J.A.K., De Moor, B.L.R.: Systematic benchmarking of microarray data classification: assessing the role of non-linearity and dimensionality reduction. *Bioinformatics* 20(17), 3185–3195 (2004)

Multiple Trend Breaks and Unit Root Hypothesis: Empirical Evidence from China's GDP(1952-2006)

Shusheng Li^{1,2} and Zhao-hui Liang¹

¹ College of Economics, Tianjin Polytechnic University, Tianjin, China

² Institute of Econometrics, Nankai University, Tianjin, China

lishusheng33@126.com, zhaohuilang@eyou.com

Abstract. Whether shocks to macroeconomic time series should be regarded as permanent or temporary has been an ongoing debate. Under the unit root hypothesis random shocks have a permanent effect on the system, and the alternative is that fluctuations are transitory. We apply the ADF test on the real GDP series of China from 1952 to 2006 by allowing for the possibility of two exogenous break points happened in 1961 and 1989 respectively. We find more evidence against the unit root hypothesis, meaning that the shocks on China's economy system are transitory except some significant events like the Great Natural Disaster in 1961 and the economy of China always grows around a stable trend path.

Keywords: Unit Root, Multiple Trend Breaks, ADF Test.

1 Introduction

During the past 2 decades there has been an ongoing debate as to whether shocks to macroeconomic time series should be regarded as permanent or temporary. The most important implication of the unit root revolution is that under unit root hypothesis random shocks have a permanent effect on the system. Fluctuations are not transitory. It runs counter to the prevailing view that business cycles are transitory fluctuations around a more or less stable trend path. It is therefore of importance to assess carefully the reliability of the unit root hypothesis as an empirical fact.

[5] applied [3] methodology to examine the 14 macroeconomic series of USA and concluded that most series are best characterized by unit root process, implying that shocks to these series are permanent. This view was challenged in [6] who rejected the unit root hypothesis for 11 of the 14 series analyzed by [5], where it was shown that a rejection of the unit root hypothesis is possible for many macroeconomic time series once allowance is made for a one-time shift in the trend function. Thus, many macroeconomic time series may be better characterized as having temporary shocks fluctuating around a broken deterministic trend function. As discussed in [1] this finding may be important for the following reasons. First, it offers an alternative picture of the persistence in macroeconomics series. Second, this approach can provide a parsimonious model for a slowly changing trend component that may be useful as a data description. Third, the implications for inference in more complex models are very different.

A key assumption of the framework proposed by [6] is that the break date of the trend function is fixed (exogenous) and chosen independently of the data. This assumption has drawn much criticism in subsequent papers, based on the argument that break dates are often chosen after looking at the data, leaving room for data-mining. This criticism was first pointed out by [2]. Several following studies have proposed procedures that address the choice of break date. These include [1],[7],[8], and [9].The strategy used in all four studies was to endogenize the choice of the break date by making it dependent. Three approaches in endogenizing the choice of the break point have been considered and all require estimation of a Dickey-Fuller (DF) type regression at all allowable break dates.

[4] extends the endogenous break methodology to allow for a two-break alternative and reexamine the unit-root hypothesis for the[5] data. They find more evidence against the unit-root hypothesis than [9], but less than [6]. Results illustrate the need for tests that are robust to misspecification with respect to the number of structural breaks. However, [8] argued that these dates can be regarded as independent of the data. First, the dates used in the previous study were chosen ex-ante and not modified ex-post. Secondly, these dates are related to exogenous events for which economic theory would suggest the effects that actually happened; In the sense described above the choice of the dates can be viewed as uncorrelated with the data. So we still believe that the assumption about the exogeneity of the choice of the break points is a good first approximation to the data.

This paper extends [6] the exogenous break methodology to a two-break alternative and examine the real GDP series of China from 1952 to 2006. The rest of the paper is organized as follows. The model and some preliminary theoretical results are presented in Section 2. Section 3 presents an empirical application and Section 4 briefly concludes.

2 Unit Root Test with Two-Break Alternative Hypothesis

2.1 One-Break Alternative Hypothesis

A given series $\{y_t\}_1^T$ is a realization of a time series process characterized by the presence of a unit root and possibly a nonzero drift. However, the approach is generalized to allow a one-time change in the structure occurring at a time T_1 ($1 < T_1 < T$). The alternative hypothesis is that the series is stationary about a deterministic time trend with an exogenous change in the trend function at time. Three different models are considered under the null hypothesis: one that permits an exogenous change in the level of the series, one that permits an exogenous change in the rate of growth, and one that allows both change. Following the notation of [6] the unit-root null hypotheses are

$$y_t = \mu + dD(T_1)_t + y_{t-1} + e_t \tag{1}$$

$$y_t = \mu_1 + y_{t-1} + (\mu_2 - \mu_1)DU_t + e_t \tag{2}$$

$$y_t = \mu_1 + y_{t-1} + dD(T_1)_t + (\mu_2 - \mu_1)DU_t + e_t \tag{3}$$

where $D(T_1)_t = 1$ if $t = T_1 + 1$, 0 otherwise; $DU_t = 1$ if $t > T_1$, 0 otherwise; $A(L) e_t = B(L) v_t$, $v_t = \text{iid}(0, \sigma^2)$, with $A(L)$ and $B(L)$ p th and q th order polynomials in the lag operator. Model (1) permits an exogenous change in the level of the series, Model (2)

allows an exogenous change in the rate of growth, and Model (3) admits both changes. The trend stationary alternative hypotheses considered are

$$y_t = \mu_1 + \beta t + (\mu_2 - \mu_1)DU_t + e_t \tag{4}$$

$$y_t = \mu + \beta_1 t + (\beta_2 - \beta_1)DT_t^* + e_t \tag{5}$$

$$y_t = \mu_1 + \beta_1 t + (\mu_2 - \mu_1)DU_t + (\beta_2 - \beta_1)DT_t + e_t \tag{6}$$

where $DT_t^* = t - T_1$ if $t > T_B$, 0 otherwise; $DT_t = t$ if $t > T_1$, 0 otherwise.

2.2 Two-Break Alternative Hypothesis

Allowing for two shifts T_1 and T_2 ($1 < T_1 < T_2 < T$), IO(Innovation Outlier) in the deterministic trend at distinct known dates, the model considered is

$$y_t = \mu + \beta t + \theta DU_{1t} + \gamma DT_{1t}^* + \omega DU_{2t} + \psi DT_{2t}^* + \alpha y_{t-1} + \sum_{i=1}^k c_i \Delta y_{t-i} + e_t \tag{7}$$

$DU_{1t} = 1$ if $t > T_1$, 0 otherwise; $DU_{2t} = 1$ if $t > T_2$, 0 otherwise; $DT_{1t}^* = t - T_1$ if $t > T_1$, 0 otherwise; $DT_{2t}^* = t - T_2$ if $t > T_2$, 0 otherwise; $c(L)$ is a lag polynomial of known order k and $1 - c(L)L$ has all its roots outside the unit circle; $A(L) e_t = B(L) v_t$, $v_t = iid(0, \sigma^2)$, with $A(L)$ and $B(L)$ p th and q th order polynomials in the lag operator. The null hypothesis of a unit root imposes the following restrictions on the true parameters of the model: $\alpha = 1$, and the alternative hypothesis is $\alpha < 1$. For $\theta \neq 0$, which means a mean shift at time T_1 , $\gamma \neq 0$, a trend shift at time T_1 , $\theta \neq 0$ and $\gamma \neq 0$, both mean and trend shift at time T_1 , the same as ω and ψ .

As in [1], it is convenient to define transformed regressors $Z_t = [Z_t^1, 1, (y_t - \mu_0 t), t + 1, DU_{1t+1}, DT_{1t+1}^*, DU_{2t+1}, DT_{2t+1}^*]'$, where $Z_t^1 = (\Delta y_t - \mu_0, \dots, \Delta y_{t-k+1} - \mu_0)$, and $\mu_0 = E(\Delta y_t)$, and a transformed parameter vector Ξ , so that equation (7) can be rewritten $y_t = \Xi' Z_{t+1} + e_t$. This transformation is adopted from and discussed by Sims et al. (1990). Let \Rightarrow denote weak convergence on $D[0, 1]$. The errors are a martingale difference sequence and satisfies $E(e_t | e_{t-1} \dots) = \sigma^2$, $E(|e_t|^i | e_{t-1} \dots) = \kappa_i$ ($i = 3, 4$), and $\sup_t E(|e_t|^{4+\xi} | e_{t-1} \dots) = \kappa < \infty$ for some $\xi > 0$. So $T^{1/2} \sum_{i=1}^{\lfloor T\lambda \rfloor} e_t \Rightarrow \sigma W(\lambda)$, for $\lambda \in [0, 1]$, W is Brownian motion. $T^{-1} \sum_{i=1}^T Z_{t-1}^1 Z_{t-1}^{1'} \xrightarrow{P} \Omega_k$, $T^{-1/2} \sum_{i=1}^T Z_{t-1}^1 e_t \Rightarrow \sigma B(1)$, $T^{-3/2} \sum_{i=1}^T Z_{t-1}^1 y_t \Rightarrow 0$, where Ω_k is a nonrandom positive semi definite matrix, and $B(1)$ is a k -dimensional Brownian motion with covariance matrix Ω_k independent of W .

Define δ_1 and δ_2 as the fractions of the sample at which the first and second breaks, respectively, occur, that is, $\delta_1 = T_1/T$ and $\delta_2 = T_2/T$. Because elements of Ξ converge at different rates, define the scaling matrix $Y_T = \text{diag}(T^{1/2}I_k, T^{1/2}, T, T^{3/2}, T^{1/2}, T^{3/2}, T^{3/2})$, define $\Gamma_T(\delta_1, \delta_2) = Y_T^{-1} \sum_{i=1}^T Z_{t-1}([T\delta_1], [T\delta_2]) Z_{t-1}([T\delta_1], [T\delta_2])' Y_T^{-1}$ and $\Psi_T(\delta_1, \delta_2) = Y_T^{-1}(\delta_1, \delta_2) \sum_{i=1}^T Z_{t-1}([T\delta_1], [T\delta_2]) e_t$. Considering the unit-root hypothesis that $\alpha = 0$. The test statistic of interest is the t -statistic associated with this hypothesis. Its asymptotic distribution can be stated as $t(\delta_1, \delta_2) \Rightarrow \int_0^1 W^*(s) dW(s) / [\int_0^1 W^*(s)^2 ds]^{1/2}$ where W^* is the continuous-time residual from a projection of a Brownian motion onto the functions $[1, s, \mathbf{1}(s > \delta_1), \mathbf{1}(s > \delta_2), (s - \delta_1)\mathbf{1}(s - \delta_1), (s - \delta_1)\mathbf{1}(s > \delta_2)]$.

Finally, we rule out the possibility that the two breaks occurred on consecutive dates. That is, we do not consider a positive shock followed by a negative shock (or vice versa) as being two separate episodes, which means $T_2 \neq T_1 + 1$. The printing area is 122 mm \times 193 mm. The text should be justified to occupy the full line width, so that the right margin is not ragged, with words hyphenated as appropriate. Please fill pages so that the length of the text is no less than 180 mm, if possible.

3 Empirical Research

3.1 Data

We apply the tests developed above to the real GDP series of China from 1952 to 2006. The real GDP series of USA were analyzed by [5], [6],[7] and [4]. We use retail price index as GDP deflator to get real GDP and take natural logarithm on the real GDP to get LNGDP series.

In figure1 the dotted line shows the plot of LNGDP. Features of LNGDP are the marked decrease between 1960 and 1962 and higher growth rate after 1989. Between 1960 and 1962, China experienced the Great Natural Disaster, so the GDP fallen dramatically. After 1989, the Chinese government adopted persistent accelerating economic development policy, which really accelerated the growth rate of GDP. Apart from these changes, the trend appears fairly stable (same slope) over the period. So we will try to examine a two trend breaks model with mean shift in 1961 and trend change in 1989. The solid line is the estimated trend line from a regression on a constant, a trend, the dummy variable $DU1_t$ taking a value of 0 prior and at 1960 and value 1 afterwards and the dummy variable $DT2_t^*$ taking a value 0 prior and at 1988 and $t - 37$ afterwards.

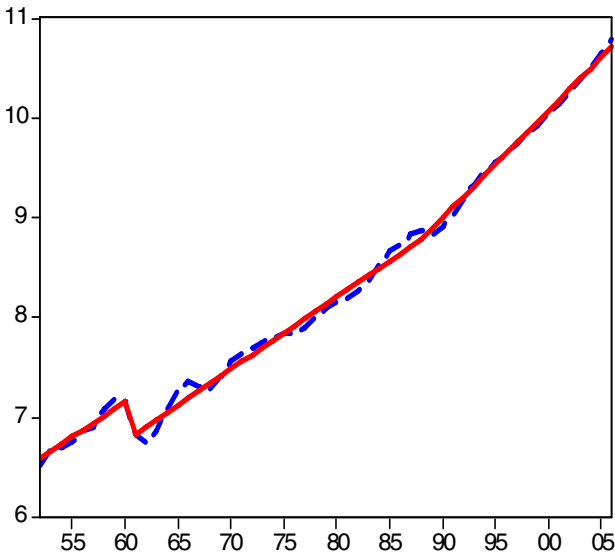


Fig. 1. LNGDP with Fitted Line

Note: The broken straight line is a fitted trend (by OLS) of the form $LNGDP = \hat{\mu} + \hat{\beta}t + \hat{\gamma}DU_1 + \hat{\psi}DT_2^*$, where $DU_t = 1$ if $t > 1960$, 0 otherwise; $DT_2^* = t - 37$ if $t > 1988$, 0 otherwise

3.2 ADF Test

Table 1 presents the results from estimating (by OLS) a regression of the augmented Dickey-Fuller type.

$$LNGDP_t = \mu + \beta t + \alpha LNGDP_{t-1} + \sum_{i=1}^k c_i \Delta LNGDP_{t-i} + e_t \tag{8}$$

Since considerable evidence exists that data-dependent methods to select the value of the truncation lag k are superior to choosing a fixed k a priori, we follow Zivot and Andrews (1992) and use the procedure suggested by [7] Start with an upper bound k_{mas} for k . If the last included lag is significant, choose $k = k_{max}$. If not, reduce k by 1 until the last lag becomes significant. If no lags are significant, set $k = 0$. We set $k_{mas} = 8$.

The first row presents the full sample regression. The coefficient on the lag dependent variable is 0.955 with t statistic for the hypothesis that $a = 1$ of -0.83. Using the critical values tabulated by Dickey and Fuller, we cannot reject the null hypothesis of a unit root. When the sample is split in three (pre-1961, 1961 to 1988, post-1988), the estimated value of a decreases dramatically. The t statistics are small enough to reject the hypothesis that $a = 1$, even at the 1 % level, for 1961-1988 and 1989-2006 sample.

Table 1. ADF Regression Analysis For LNGDP

Regression: $LNGDP_t = \mu + t + LNGDP_{t-1} + \sum_{i=1}^k c_i \Delta LNGDP_{t-i} + e_t$					
Period	k	μ			P
1952-2006	4	0.30 (0.92)	0.01 (1.40)	0.955 (-0.83)	0.9558
1952-1960	0	8.06 (2.65*)	0.10 (2.40)	-0.23 (-2.61)	0.2899
1961-1988	1	4.81 (7.91**)	0.05 (8.03**)	0.30 (-7.89**)	0.0000
1989-2006	1	6.25 (7.46**)	0.08 (6.42**)	0.29 (-6.31**)	0.0004
1961-2006	4	1.03 (6.36**)	0.01 (3.01**)	0.85 (-2.48)	0.3365

Note: * means significant at 5% level, ** means 1% level

Two features are worth emphasizing from this example: (a) the full sample estimate of a is markedly superior to any of the split sample estimates and relatively close to one. It appears that the 1961 and 1989 trend breaks are responsible for the near unit root value of a ; and (b) even though the split sample 1961-1988 and 1989-2006 respectively are powerful enough to reject the hypothesis that $a = 1$ even at 1% level, if we combine them together 1961-2006, we fail to reject the null hypotheses even at 10% level. It would be useful, in this light, to have a more powerful procedure based on the full sample that would allow the 1961 and 1989 break to be exogenous.

3.3 Two-Break Unit Root Test

Model (7) allows for two breaks in both the intercept and the slope of the trend function. The results for this model reported in table 2. The k chosen criteria is the same as that of model(8).

Both θ and ψ are significant at 1% level, while γ and ω are insignificant, which mean GDP of China dramatically fallen down in 1961 and had a significant high growth rate after 1989, which are in line with our judgment above. t_α is only -9.74 , much less than -7.76 (the critical value of 1%), so we can reject unit root hypothesis at 1% level. The result is the same as [8] about the real GNP of USA from the first quarter 1947 to the third quarter 1986. The real GDP series of China from 1952 to 2006 is a trend stationary process with two trend breaks respectively in 1961 and 1989, not a stochastic trend process. The underlying idea is that although the central government makes complex fiscal and monetary policies to stipulate the economy, the shocks on the economy system are transitory except some significant events like the Great Natural Disaster in 1961 and persistently accelerating policy after 1989. The economy of China always grows around a more stable trend path.

Table 2. Two-break Unit Root Test

$$y_t = \mu + \beta t + \theta DU1_t + \gamma DT1_t^* + \omega DU2_t + \psi DT2_t^* + \alpha y_{t-1} + \sum_{i=1}^k c_i \Delta y_{t-i} + e_t$$

	μ	β	θ	γ
	4.71 (9.76**)	0.06 (5.92*)	-0.30 (-6.54**)	-0.03 (-1.24)
Critical Value	1% -5.73	1% -6.24	1% -5.18	1% -6.15
	5% -4.40	5% -5.26	5% -4.34	5% -5.14
	10% -3.65	10% -4.58	10% -3.85	10% -4.55
	50% -0.04	50% 0.04	50% -0.41	50% -0.12
	90% 3.65	90% 4.47	90% 3.85	90% 4.52
	95% 4.43	95% 5.15	95% 4.34	95% 5.13
	99% 5.52	99% 6.16	99% 5.19	99% 6.15
	ω	ψ	α	k
	-0.01 (-1.23)	0.03 (7.63**)	0.26 (-9.74**)	1
Critical Value	1% -5.27	1% -6.30	1% -7.76	
	5% -4.41	5% -5.31	5% -7.03	
	10% -3.92	10% -4.69	10% -6.71	
	50% 0.08	50% 0.02	50% -5.68	
	90% 3.92	90% 4.74	90% -4.78	
	95% 4.43	95% 5.37	95% -4.37	
	99% 5.33	99% 6.27	99% -4.14	

Notes: The critical values taken from Luan(2005) computed using 60 observations and 5000 replications, * means significant at 5% level, ** means 1% level

4 Conclusion

This paper has attempted to resume debate regarding the relationship between the unit root hypothesis and structural breaks. We have extended the exogenous one break

model to the case of two breaks. In particular, we have used real LNGDP of China from 1952 to 2006 as an example to illustrate that inference related to unit roots is sensitive to the number of assumed breaks. We have shown that the results obtained without trend break or using one break are reversed with two breaks. The two breaks we choose are in 1961 and 1989 by exploiting the LNGDP plot, and we explain why there is a trend break. Between 1960 and 1962, China experienced the Great Natural Disaster, so the GDP fallen dramatically. After 1989, the Chinese government adopted persistent accelerating economic development policy, which really stipulated the growth rate of GDP. The model shows that there is a mean shift in 1961 and a trend shift in 1989, which are in line with our judgment and the real world. So the real GDP series of China from 1952 to 2006 is a trend stationary process with two breaks, not a stochastic trend process, which has significant economic policy meaning that the shocks on China's economy system are always transitory except some significant events and the economy of China always grows around a stable trend path.

References

1. Banerjee, A., Lumsdaine, R.L., Stock, J.H.: Recursive and Sequential Tests of the Unit-Root and Trend-break Hypotheses: Theory and International Evidence. *Journal of Business and Economic Statistics* 10, 271–288 (1992)
2. Christiano, L.J.: Searching for a Break in GNP. *Journal of Business and Economic Statistics* 10, 237–250 (1992)
3. Dickey, D.A., Fuller, W.A.: Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association* 74, 427–431 (1979)
4. Lumsdaine, R.L., Papell, D.H.: Multiple Trend Breaks and the Unit-root Hypothesis. *The Review of Economics and Statistics* 79, 212–218 (1997)
5. Nelson, C.R., Plosser, C.I.: Trends and Random Walks in Macro-economic Time Series: Some Evidence and Implications. *Journal of Monetary Economics* 10, 139–162 (1982)
6. Perron, P.: The Great Crash, the Oil Price Shock and the Unit Root Hypothesis. *Econometrica* 57, 1361–1401 (1989)
7. Perron, P., Vogelsang, T.J.: Testing for a Unit Root in a Time Series with a Changing Mean: Corrections and Extensions. *Journal of Business & Economic Statistics* 10, 467–470 (1992)
8. Perron, P.: Further Evidence on Breaking Trend Functions in Macroeconomic Variables. *Journal of Econometrics* 80, 355–385 (1997)
9. Zivot, E., Andrews, D.W.K.: Further Evidence on the Great Crash, the Oil-price Shock, and the Unit-root Hypothesis. *Journal of Business and Economic Statistics* 10, 251–270 (1992)

An Adaptive Wavelet Networks Algorithm for Prediction of Gas Delay Outburst

Xinyu Li

The Second Coal Mine, Pingdingshan Coal Co. Ltd, Pingdingshan 467000, China

Abstract. An adaptive wavelet networks algorithm for prediction of gas outburst is proposed in this paper. First, adaptive clustering algorithm is first used to determine initial parameters of wavelet network according to the results of the clustering. Then genetic algorithm and SVM-RFE is adopted to tune the structure of the wavelet network and adjust the network parameters to improve generalization performance. Finally, the simulation for prediction of gas outburst is discussed, and the result shows the validity of the proposed algorithm.

Keywords: Wavelet, Prediction, Gas delay outburst.

1 Introduction

Coal and gas delay outburst often takes place sometime after the operations or explosions. In general, the abnormal events, such as the continued increase of gas emission, the increasing of gas concentrations, increased pressure on tunnel roof [1], will appear before delay outburst. However, the relationship between characteristics and categories of gas delay outburst is seriously nonlinear [2]. So the key to monitoring gas delay outburst is to select a appropriate classifier. The most existing methods for prediction of gas delay outburst adopt the BP neural network [3]. However, BP neural network is difficult to select parameters, neuron nodes in hidden layer etc., seriously affecting the performance of BP. To improve the performance of neural network, people propose many types of neural networks, such as RBF neural network, wavelet network, fuzzy neural networks [4]. Although the wavelet basis with excellent local time-frequency characteristics and multi-scale analysis inherits from wavelet analysis adapts to describe short-term, high-frequency signal and the non-uniform sampling data [5,6], how to select parameters of the radial wavelet-based network as well as the structure of network and training is a challenging task. Most methods determine the parameters of wavelet basis in the framework of wavelet according to distribution of samples. Then AIC (Alaikes information Criterion) integrating with other algorithms, such as genetic algorithms, Gram-Schmit algorithm, the gradient descent etc., adjust the parameters of WN [7]. However, there are some problems exist in these methods, such as improper initial parameters, long training time, learning criteria based on experience risk minimization.

Considering the above mentioned problems, this paper an adaptive WN classification algorithm is presented. Firstly, adaptive clustering algorithm is applied to cluster data samples; Secondly, it sets the parameters of wavelet basis initialized accordance to the cluster radius and the cluster centers in term of the framework of wavelet.

Genetic algorithm (GA) and Support vector machine for recursive feature elimination (SVM-RFE) try to find optimal or suboptimal parameters and the structure of WN. The method can make full use of Support vector machine (SVM) and achieved high generality. Choosing 8 features from the real data, it establishes 3 classification models being similar to the SVM multi-category classification strategy called ‘one against all’. The result of simulation shows that the algorithm is can achieve higher precision than the existing algorithms for prediction gas outburst.

2 Related Knowledge

According to the wavelet analysis theory [5], wavelet functions have limited supports in the time and frequency. Define the extent of the any given error as $\mathcal{E} > 0$, any functions $f(\mathbf{x}) \in L^2(R^d)$ can almost be reproduced in finite subset of time-frequency, that is

$$f(\mathbf{x}) \cong \sum_{i=1}^{n_a} \sum_{j=1}^{n_i} w_{ij} \psi_{a_i, \mathbf{b}_{ij}}(\mathbf{x}) + \bar{b} = \sum_{i=0}^{n_a} f_i(\mathbf{x}) + \bar{b} \tag{1}$$

where, $\psi_{a_i, \mathbf{b}_{ij}}(\mathbf{x}) = 1/\sqrt{a_i} \psi((\mathbf{x} - \mathbf{b}_{ij})/a_i)$ the basic wavelet; \mathbf{b}_{ij} shift factor; a_i scale factor; w_{ij} the wavelet coefficients $i = 1, 2, \dots, n_a$, $j = 1, 2, \dots, n_i$; n_a the number of scales; n_i the number of the wavelet function in accordance with scale i ; \bar{b} offset and a constant number; f_i at scale i can be approximated by

$$f_i(\mathbf{x}) = \sum_{j=1}^{n_i} w_{ij} \psi_{a_i, \mathbf{b}_{ij}}(\mathbf{x}) \tag{2}$$

From the above equation, wavelet can be taken as basic function to approximate any function of space $L_2(R^d)$. Obviously, the above mentioned formula can be characterized by wavelet neural network. It approximates functions in several scales. The WN is superiority to SVM in the condition of appropriate structure of WN and suitable parameters.

3 The Dynamic Adaptive Clustering Algorithm

Radial basis wavelet $\varphi_{\sigma, \mathbf{c}}(\mathbf{x}, \mathbf{c}) = (n - \|\mathbf{x} - \mathbf{c}\|^2 / \sigma^2) \exp(-\|\mathbf{x} - \mathbf{c}\|^2 / \sigma^2)$ is similar to function of RBF. σ the width factor; \mathbf{c} is the center factor. Since the structure of WN is similar to that of RBF neural network, we can use the clustering algorithm to determine the initial parameters of radial basis wavelet. The results of clustering determine whether the initial parameters are suitable or not, directly influencing learning algorithm’s speed. In view of above discussion, this article uses dynamic adaptive

algorithm mentioned in ref [8]. The clustering algorithm is a supervised clustering and can reduce the uncertainty. The leaning criteria is as follows:

$$J = \sum_{i=1}^k \sum_{\mathbf{x} \in \Gamma_i} \|\mathbf{x} - \mathbf{m}_i\|^2 \tag{3}$$

Here, Γ_i is the sample set belonged to the class i ; $|\Gamma_i|$ is the number of samples belonged to the class i ; \mathbf{m}_i is mean of category i , $\mathbf{m}_i = \frac{1}{|\Gamma_i|} \sum_{\mathbf{x}_j \in \Gamma_i} \mathbf{x}_j$. The idea of dynamic adaptive algorithm is to select cluster center randomly. If the distance between the sample and center of any cluster is larger than the threshold, then take the sample with maximum distance as a new cluster. Finally, each sample will close to the center of a cluster with the shortest distance. The algorithm can be referred ref [8]. It needs to test the cluster results after clustering. The testing method is to move a sample from one cluster to another cluster and calculate the difference of mean of clusters; Then judge the cluster's new sum of square of errors whether smaller or not. If sum of square of errors become larger, the clustering can stop, otherwise continue on.

4 Adaptive Training Algorithm for WN

4.1 Initial Structure and Parameters of WN

It is assumed that input and output variables of the sample are in $[0, 1]$, otherwise, it requires to normal. For wavelet with finite support, given a scale, it is easy to determine the parameters [9]. However, most of the radial basis of wavelet is not tight support, but can rapidly decline to zero. In view of this point, we can remove wavelet nodes in some data samples to be semi-tightly supported.

For m -dimensions data $0 \leq x_i \leq 1, i = 1, 2, \dots, m$, support domain of radial basis wavelet is $x_i \in [-3, 3], i = 1, 2, \dots, m$, let the center be $\mathbf{c} = [c_1, c_2, \dots, c_m]^T$, radial basis wavelet function is $\varphi_{\sigma, \mathbf{c}}(\mathbf{x}, \mathbf{c}) = \left(n - \|\mathbf{x} - \mathbf{c}\|^2 / \sigma^2 \right) \exp\left(-\|\mathbf{x} - \mathbf{c}\|^2 / \sigma^2 \right)$. Given scale factor σ , wavelet basis meets the conditions in tight support domain: $|x_i - c_i| / \sigma \leq 3$, that is $-3\sigma + x_i \leq c_i \leq 3\sigma + x_i, i = 1, 2, \dots, m$. Given the cluster center and the cluster radius, from $|x_i - c_i| / \sigma \leq 3$, we can deduce that $\sigma \geq \frac{r}{3}$. Thus, for any given cluster, we can determine the number of wavelet and scale factor (or width- factor), translation factor according to cluster information.

According to experience, the requirement for the support border of wavelet are not strict, in other words, the selection of translation factors of wavelet functions is robustness. However, the scale factor of wavelet function influences the output of wavelet

functions significantly. Therefore, this article creates a certain number of wavelet nodes based on the cluster. Scale and translation factors of each wavelet node generates randomly in the following range:

$$\sigma'_i \in \left[(1-10\%)*\frac{r_i}{3}, (1+10\%)*\frac{r_i}{3} \right] \tag{4}$$

$$c'_i \in \left[c_i*(1-20\%), c_i*(1+20\%) \right] \tag{5}$$

Obviously, there is some redundant wavelet nodes in the framework of wavelet, and parameters of wavelet nodes may not be accurate. It needs to eliminate redundant wavelet neurons and adjust the parameters of wavelet basis. This paper adopts genetic algorithms and SVM-RFE to train WN and improve the performance of WN.

Assuming that there are M candidates of wavelet neurons in WN, WN can be converted into the following linear form (linear-in-parameters) after determining the parameters of wavelet neurons:

$$y(\mathbf{x}) = \sum_{i=1}^M w_i \varphi_i(\mathbf{x}) + e \tag{6}$$

The linear model includes some unimportant neurons and inappropriate parameters of wavelet elements, leading to generate the problem of low performance. Therefore, it is necessary to eliminate the redundant neurons. If regard each neuron as a feature of input data, the problem of eliminating the redundant wavelet neurons can be translate into feature selection. With good generality, SVM has widely used to feature selection [10, 11]. The article adopts the SVM-REF (Recursive feature selection based on support vector machine) proposed by Ref [10] to eliminate the redundant wavelet nodes. Since the adjustment of parameters of WN is a complex non-linear problem, genetic algorithms (GA) is applied to optimize parameters of Wavelet nodes.

4.2 Training Algorithm of WN Based on Genetic Algorithm and SVM-RFE

A large number of experiments show that, the GA can be used to obtain the optimal or near optimal neural network parameters. In order to increase the speed of training algorithm, training algorithm of WN is put forward. The method optimizes parameters of wavelet nodes and the weights of WN separately. SVM-RFE algorithm determines the parameters and network topologies while GA adjusts the parameters of wavelet, the method is shown in Fig.1. σ_i, c_i, f_i scale factor, translation factor and the fitness value of each individual i_{th} , and fitness value is the test accuracy of individual. The upper layer receives wavelet neuron parameters which are optimized through GA, SVM-RFE is used to calculate output weights, the importance of wavelet neuron is sorted, then the unimportant wavelet neurons is eliminated.

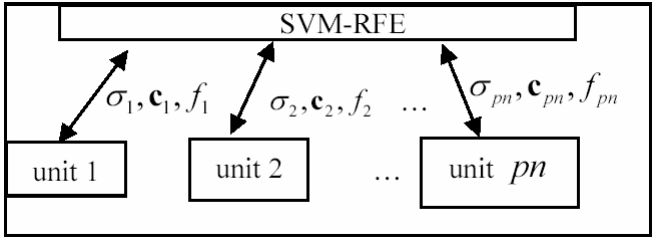


Fig. 1. Learning algorithm of WN

SVM-RFE selects features recursively. It improves the accuracy discrimination by removing the irrelevant features and retaining feature subsets with higher importance. In SVM-RFE, SVM is used as a classifier. The cost function of SVM-REF is:

$$J(\alpha) = \frac{1}{2} \alpha^T H \alpha - \alpha^T \mathbf{1} \tag{7}$$

Where $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_M]$; $H = [y_i y_j \mathbf{p}_i^T \mathbf{p}_j]_{i,j=1,2,\dots,N}$; N the number of samples; $\mathbf{p}_j = [\phi_1(\mathbf{x}_j), \dots, \phi_M(\mathbf{x}_j)]^T$; $\mathbf{1}$ a $M \times 1$ vector that all the elements are 1. For the feature J , the importance is defined as follows:

$$\Delta J(i) = J(\alpha) - J(\alpha - i) \tag{8}$$

The SVM-RFE algorithm, and corresponding issues can be seen in literature [12].

5 Simulation

Generally, there are some certain phenomena before the happening of gas delay outburst, such as continued increase of gas concentrations, and so on. The phenomena have some similarities and differences [3]. In addition to the features proposed in Ref [3], such as the peak of underground gas emission, increased gradient of gas, transfinite time of gas, decline in gas gradient, the paper makes use of wavelet analysis to extract features in Ref [2]. Db4 wavelet is used to divide the signal into the 3-scale and the noises can be removed through soft-threshold. Then, we take square mean of wavelet coefficients in 3-scale as 4 features in the frequency domain.

According to gas features, we can predict delay gas outburst after the guns, probe and check sensors of gas, failure in communication line. For simplicity, we have to divide monitoring categories into state with troubles, normal state after guns, predicting state of outburst (that is, the methane contained in gas is more than 3% which sustained for more than 10s). Since the adaptive WN can only classify two classification samples, "one-against-all" strategy is adopted to train multi-classifier. Therefore, it is needed to train 3 WNs, the output is ± 1 which are positive class and negative

class (that is, class 1 and class 2). The positive classes of WN1, WN2, WN3 are state with failure, normal state after guns, predicting state of outburst while the negative categories are corresponding to classes except the normal.

In training algorithm of WN, punishing factor of SVM is decided according to punish parameters 5-fold cross-validation and is unchanged in order to reduce the amount of training. The initial number of WN wavelet neuron is no more than 40, the parameters of genetic algorithm are achieved according to the recommended value. Each number of individuals is 30, and the evolution of algebra is no more than 30.

190 samples of data are generated from a group data of mine gas outburst, normal data after blasting, as well as data of colliery tragedy in Shanxi. Among those, the delayed gas outburst samples are 66; the normal data samples are 96; failure data samples are 28. Selecting randomly 1 / 2 from them as training sample, and the remaining samples are for testing sample. Matlab7.0 is used as simulation language and its own in-house optimization and genetic algorithms are used. SVM code is the source code from <http://alex.smola.org/code.html>. During the process of training Wavelet network, the error of training reduces along with the genetic algorithm increases escaping from increasing margin of error in the BP neural network training process. Each is tested through WN1, WN2, WN3 testing. The use of the "one-against-all" strategy determines the final sample of the new category. The accuracy rate of test model is up to 85.2 percent by using all the test-sample.

6 Conclusion

A new learning algorithm for WN is presented. SVM-REF is used to determine the structure of WN while GA adjusts WN parameters. the method is based on structure risk minimization and initial parameters are determined properly by dynamic adaptive clustering algorithm, which guarantee generality of WN and short training time. The simulation results show that this method has a higher classification of the superiority than existing gas outburst forecast algorithm and overcomes the shortcomings of existing WN algorithms.

Acknowledgments. The author would express appreciation for the financial support of the China Planned Projects for Postdoctoral Research Funds under Grant NO.20060390277 and the Jiangsu Planned Projects for Postdoctoral Research Funds under Grant NO.0502010B. The author also would express thanks for Six Calling Person with Ability Pinnacle under Grant NO.06-E-052, Jiangsu Technology Projects Research Funds under Grant NO.BG2007013, and Jiangsu Graduate Training and Innovation Project.

References

1. Yu, Q.X.: Mine Gas Prevention and Control. China University of Mining and Technology Press, Xuzhou (1992)
2. Dang, J.J., Li, X.J., Zhu, R.Q.: Study on Abstracting Signal of Coal and Gas Outburst Prediction Based on Wavelet Analysis. *Zhongzhou Coal* 151, 6–7 (2008)

3. Luo, X.R., Yang, F., Kan, Y.T., Zhang, A.R.: Research on Real-Time Alarm Theory of Delayed Coal and Gas Outburst. *Journal of China University of Mining and Technology* 37, 163–166 (2008)
4. Fan, C.N., Li, K.K., Chan, W.L., Yu, W.Y., Zhang, Z.N.: Application of Wavelet Fuzzy Neural Network in Locating Single Line to Ground Fault (SLG) in Distribution Lines. *Electrical Power and Energy Systems* 29, 497–503 (2007)
5. Yang, F.S.: *Wavelet Analysis and Application of Engineering*. Science Press, Beijing (1999)
6. Eric, A., Rying, Griff, L.B., Jye-Chyi, L.: Focused Local Learning with Wavelet Neural Networks. *IEEE Trans. on Neural Netw.* 13, 304–320 (2002)
7. Lv, L.H.: *Complex Industrial Systems Based on Wavelet Network and Robust Estimation Modeling Study*. Zhejiang University (2001)
8. Xiao, D., Lin, J.G.: The RBF Neural Network Faces Recognition Based on Dynamic Clustering. In: *Chinese Control and Decision Conference (CCDC 2008)*, pp. 3996–3999 (2008)
9. Kugarajah, T., Zhang, Q.: Multidimensional Wavelet Frames. *IEEE Trans. Neural Netw.* 6, 1552–1556 (1995)
10. Guyo, I., Weston, J.: Gene Selection for Cancer Classification Using Support Vector Machines. *Machine Learning* 46, 389–422 (2002)
11. Gerda, C., Christophe, C., Johan, V.K.: An Information Criterion for Variable Selection in Support Vector Machines. *Journal of Machine Learning Research* 9, 541–558 (2008)
12. Mao, Y.: *The Research and Application of Feature Selection Methods Based on SVM* PhD Thesis. Zhejiang University (2006)

Traffic Condition Recognition of Probability Neural Network Based on Floating Car Data

Gengqi Guo^{1,2}, Chengtao Cao^{1,2}, Jiuzhong Li³, and Shuo Shi³

¹ South China University of Technology, Guangzhou 510640, China

² Guangdong Communication Polytechnic, Guangzhou 510650, China

³ Guangdong Industry Technical College, Guangzhou 510300, China

ggq@gdcp.cn, jncct@163.com, gzlijz@163.com, shi62@163.com

Abstract. A traffic condition recognition method based on floating car data was proposed by analyzing Probability Neural Network (PNN) and Global K-means algorithm. The related factors of traffic condition and the collection method of floating car data were presented. A probability neural network classifier was designed using Global K-means algorithm and applied to the recognition of traffic condition with floating car data. The experiment results showed that the method could recognize traffic condition well. The accurate rate is satisfactory.

Keywords: Floating car, Traffic condition, Probability neural network, Traffic information, Global K-means.

1 Introduction

With the extensive application of the GPS, GIS and wireless communications, traffic information collection from floating cars equipped GPS devices has become a new way to collect traffic information [1]. How to recognize traffic condition according to the collected traffic information is very important for the car navigation and the road supervisions [2,3]. By using Global K-means algorithm to design probability neural network, this paper proposed a traffic condition recognition method based on floating car data. The experiment showed that the method could recognize traffic condition accurately.

Because the ITS common information platform uses taxi as floating car and taxi will have two situations which are no-load and load, some floating car data can't reflect traffic condition truly. We must process the floating car data according characters of taxi.

(1) Data processing when taxi is load

The taxi has two situations: load and no-load. When it is load, the driver will reach the destination with the speed. Under this situation, the floating car data can reflect traffic condition accurately. When it is no-load, the driver will find the guest that he will drive slowly even stop to wait for the guest. The floating data under this situation can't reflect traffic condition. Because every taxi is equipped the charge machine, when it is load the charge machine is load, when it is no-load the charge machine is

no-load. The status of the charge machine can be delivered to the supervision center. So we can judge the situation of taxi according that of charge machine. If the charge machine is load, the floating car data is effective. The place of the taxi can be detected by GPS and GIS. When the charge machine is no-load but the taxi is in the high way and other place that forbidden to pick guest, the floating car data is effective too.

(2)Data processing when taxi stops

When the taxi stops, there are two situations: first, the traffic signal light is red or the traffic congestion; the floating car data is effective and can reflect traffic condition accurately. Second, the taxi stops for waiting guest; the floating car data is invalid. The data can be processed by judging the status of charge machine when taxi stops. If the charge machine is load when the taxi stops, the floating car data is effective. If the charge machine is no-load when the taxi stops but the taxi is in the highway and other place that forbidden to pick guest, the floating car data is effective too.

2 Traffic Information Calculation Based on Floating Car Data

The recognition of traffic condition needs concrete traffic information. The traffic information used in common includes vehicle flow data, speed, occupation rate, road segment traveling time and queue length etc. Because there is certain relativity between above traffic parameters and the floating car can detect several parameters, it's important to choose representative traffic parameters which the floating car can detect and can reflect the urban traffic condition [2].

Floating car equipped GPS device can detect the traffic information including latitude and longitude, speed, road segment traveling time and direction [3]. Among them, speed and road segment traveling time are closely related with traffic condition. Because the instant speed is changing, it may cause deviation. Unless the abrupt affairs, the urban traffic condition has seldom abrupt change in a short time. So the average speed can reflect the traffic condition better than the instant speed. This paper chooses road segment traveling time, average speed and the change rate of the instant speed to recognize the traffic condition.

(1)Average speed.The floating car can detect the instant speed at any time frequently. But if the sample cycle is too short the correspondence cost will increase. Traffic condition recognition needs short time average speed of the target road segment. Supposed that the floating car can detect traffic data of two points on the target road segment at least.,then the instant speed of the the floating car is:

$$V_i = \frac{L}{\Delta t} \quad (1)$$

where L is the distance between the two neighbor sample points, Δt is the time partition between two sample points.

If we get the first GPS point and the last GPS point on the target road, the average speed of the whole target road can be calculated approximatively by the following equation:

$$\bar{V}_i = \frac{\bar{L}}{\Delta t}$$

where \bar{L} is the distance between the first GPS point and the last GPS point., Δt is the the time partition between these two sample points.

(2)Target road traveling time. The data that floating car detected including vehicle’s position and time. Supposed that the whole length of the target road segment is \bar{S} , the length and the time between the first GPS point and the last GPS point on the target road is \bar{L} and Δt respectively. The target road traveling time \bar{T} is deduced approximatively by:

$$\bar{T} = \frac{\bar{S}}{\bar{V}_i} = \frac{\bar{S}}{\bar{L} / \Delta t} = \frac{\bar{S} \Delta t}{\bar{L}} \tag{2}$$

(3)The change rate of the instant speed. If the instant speed of the floating car is steady, the interference of other vehicles to the car is small and the traffic condition is well. The change rate of the instant speed is defined by

$$C_i = \frac{\nabla V_i}{\bar{V}_i} = \frac{\sqrt{\frac{1}{m} \sum_{j=1}^m (V_{ij} - \bar{V}_i)^2}}{\bar{V}_i} \tag{3}$$

where V_{ij} is the jth instant speed in the instant speed sequence of ith floating car.

3 Related Factors Analysis of Traffic Condition

Depicted in Fig.1.Supposed P_0 is the last GPS location before it reaches the target road. The length between P_0 and the first GPS location P_1 on the target road is S_0 , its average speed is V_0 , its average traveling time is T_0 ; the length between the first GPS location P_1 and the last GPS location P_2 on the target road is S_1 , its average speed is V_1 , its average traveling time is T_1 ; the length between P_2 and the first GPS location P_3 on the next road segment is S_2 , its average speed is V_2 , its average traveling time is T_2 . Supposed that the whole length between P_0 and P_3 is S , the average speed of this whole track is \bar{V} .

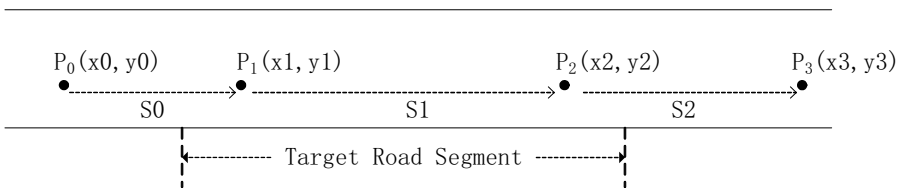


Fig. 1. Illustration of Road Segment

Because the traveling time of the whole route is equal to the sum of the traveling time of its part route [4], we can deduce the following equation:

$$\frac{S}{V} = \frac{S_0}{V_0} + \frac{S_1}{V_1} + \frac{S_2}{V_2} \tag{4}$$

Eq(4) can be converted into Eq(5):

$$\frac{1}{V} = \frac{S_1}{S} \left(\frac{S_0 V_2 + S_2 V_0}{S_1 V_0 V_2} + \frac{1}{V_1} \right) \tag{5}$$

The average speed of target road V1 can be given by the expression:

$$\begin{aligned} \frac{1}{V_1} &= \frac{S}{S_1 V} - \frac{S_0 V_2 + S_2 V_0}{S_1 V_0 V_2} = \frac{1}{S_1} \left(\frac{S}{V} - \frac{S_0}{V_0} - \frac{S_2}{V_2} \right) \\ \frac{1}{V_1} - \frac{1}{V} &= \frac{S - S_1}{S_1} \cdot \frac{1}{V} - \frac{1}{S_1} \left(\frac{S_0}{V_0} + \frac{S_2}{V_2} \right) = \frac{S_0 + S_2}{S_1} \cdot \frac{1}{V} - \frac{1}{S_1} (T_0 + T_2) \end{aligned} \tag{6}$$

From Eq.(6) we can see that the error margin between average speed of target road and average speed of whole road has important relation with S_1 . The method reducing error is to enlarging the length of target road S_1 . There are three conditions:

(1) When V_0 is far larger than V_2 , S_2 is small. We have

$$\frac{1}{V_1} = \frac{S}{S_1 V} - \frac{S_0 V_2 + S_2 V_0}{S_1 V_0 V_2} \approx \frac{S}{S_1 V} - \frac{S_2}{S_1 V_2} = \frac{1}{S_1} (T - T_2) \tag{7}$$

(2)When V_0 is equal to V_2 , we have:

$$\frac{1}{V_1} = \frac{S}{S_1 V} - \frac{S_0 V_2 + S_2 V_0}{S_1 V_0 V_2} \approx \frac{S}{S_1 V} - \frac{S_0 V_2 + \alpha \cdot S_2 V_2}{\alpha \cdot S_1 V_2 V_2} = \frac{1}{S_1} \left(T - T_2 - \frac{S_0}{\alpha \cdot V_2} \right) \tag{8}$$

(3)When V_0 is far smaller than V_2 , S_0 is small. We have:

$$\frac{1}{V_1} = \frac{S}{S_1 V} - \frac{S_0 V_2 + S_2 V_0}{S_1 V_0 V_2} \approx \frac{S}{S_1 V} - \frac{S_0}{S_1 V_0} = \frac{1}{S_1} (T - T_0) \tag{9}$$

From above analysis, we can see that it needs to choose the GPS data of two places which are near but not in the target road as the departure place and the terminal place to calculate the average traveling time.

3.1 Traffic Condition of Neighbor Road Segment

According to the relation between speed, distance and time, we can compute the absolute error margin of the speed of a target road and that of whole road:

$$\begin{aligned} \frac{S_1}{T_1} - \frac{S_0 + S_1 + S_2}{T_0 + T_1 + T_2} &= \frac{S_1(T_0 + T_2) - T_1(S_0 + S_2)}{T_1(T_0 + T_1 + T_2)} \\ &= \frac{S_1}{T_0 + T_1 + T_2} \left(\frac{T_0 + T_2}{T_1} - \frac{S_0 + S_2}{S_1} \right) = \frac{T_0(V_1 - V_0) + T_2(V_1 - V_2)}{T} \end{aligned} \tag{10}$$

From Eq.(10) we can see, the error margin of target road speed and whole road average speed is directly relative with the error margin of average speeds of neighbor road segments. The affect of traffic condition of neighbor road segment to the result is showed as Table 1.

Table 1. The affect of neighbor road to the result

NO.	Traffic condition of entrance segment	Traffic condition of target segment	Traffic condition of exit segment	error
1	slow	slow	slow	small
2	slow	slow	fast	large
3	slow	fast	slow	smaller
4	slow	fast	fast	small
5	fast	fast	slow	large
6	fast	fast	fast	larger
7	fast	fast	slow	small
8	fast	fast	fast	small

From Table 1 we can see, because of the affect of neighbor road segment, considering that the change of traffic condition is slow, there will be deviation of the consistency between the traffic condition of target road and computing result, namely be partial to great, or be partial to small. It is difficult to predict the relativity of traffic condition of different direction road segment, so the error margin of this kind of method is more difficult to cancel.

3.2 Traffic Condition of Exit Road Segment

If the first sample place is moved from the entrance road segment to the target road segment, According to the relation between speed, distance and time, we can compute the absolute error margin of the speed of a target road and that of whole road:

$$\begin{aligned} \frac{S_1}{T_1} - \frac{S_1 - S_0 + S_2}{T_1 - T_0 + T_2} &= \frac{S_1(T_2 - T_0) - T_1(S_2 - S_0)}{T_1(T_1 - T_0 + T_2)} \\ &= \frac{S_1}{T_1 - T_0 + T_2} \left(\frac{T_2 - T_0}{T_1} - \frac{S_2 - S_0}{S_1} \right) = \frac{T_0 \cdot (V_1^1 - V_1) + T_2 \cdot (V_1 - V_2)}{T_1 - T_0 + T_2} \end{aligned} \tag{11}$$

where, V_1^1 is the average speed from the departure place to the sample place, which can be named ex-segment of target segment. With the similar analysis as Table 1, we can get Table 2. Because the speed of ex-segment of target segment is relative with that of target segment, it is considered that if the speed of ex-segment is slow, the speed of the target segment is slow; however if the speed of ex-segment is fast, we can't conclude that the speed of the target segment is fast. Table 2 is showed as:

Table 2. The affect of the exit road to the result

No.	Ex-target segment	Target segment	Exit segment	error
1	slow	slow	slow	small
2	slow	slow	fast	large
3	fast	slow	slow	small
4	fast	slow	fast	small
5	fast	fast	slow	small
6	fast	fast	fast	small

From Table 2, the affect of the speed of exit road segment to that of target road segment is much steady. With the increase of sample, T_0 and T_2 , S_0 and S_2 will offset. See from actual circumstance, the speed of the entrance road segment is fast generally, because of the offset of error margin, this method has good adaptability for this circumstance.

4 Urban Traffic Condition Recognition Based on PNN

According to above analysis, the following problems need to be solved in order to calculate the traffic condition of certain road segment using the floating car data [5].

1. Fixing on the departure place: Record the last GPS data of the ex-segment of target segment, including time T_1 , latitude and longitude(x1,y1), instantaneous velocity S_1 . These GPS data are used as start point of calculation.
2. Fixing on terminal place. The terminal place needs to satisfy the following conditions.
 - a. At least there are target road segment between the starting point and the terminal place.
 - b. GPS data of terminal place must be accord with condition a. Considering the floating car's actual driving characteristics, there should be special circumstances, such as determining whether the taxi is parking for waiting for customers, treatment of situations that can't position accurately.

4.1 Probability Neural Network (PNN)

Probability neural network (PNN) is a neural network model put forward by Specht in 1988. Compared with the traditional neural networks, such as BP, the main advantages of PNN contains the following points [6]:

- (1) Without BP network's error back-propagation process, its training speed is very fast, Although the training time is slightly larger than the time to read the data.
- (2) Its Convergence is good. No matter how complicate the issue of classification is, as long as enough training samples, it's sure to get the Bayesian optimal solution.
- (3) Its network structure can be designed flexibly, allowing to increase or decrease training samples without re-training for a long time.

Based on the above advantages, PNN has been widely used in pattern recognition, fault diagnosis and other fields. Due to hidden layer nodes of the conventional PNN are equal to the number of training samples, when the training sample size is large, PNN structure will become very complicated to hardware. Therefore it is necessary to carry out training set reduction to achieve the optimal design of the PNN. In this paper, Global K-means algorithm is used to select PNN hidden nodes, while its the kernel parameters are given using the experience of Method.

Traditional K-means algorithm is a means process that the samples were divided into K clusters for a given set of limited samples O and given several types of K, which making the sum of the distance of each sample with the cluster center minimum. Since the start of each category was selected for the center randomly, the traditional K-means algorithm easily converges into a local minimum.

Likas proposed Global K-means algorithm aiming at the shortcomings of traditional K-means algorithm [7]. When collected samples are less than 150 and the number of categories is smaller than 8, the best classification result is available. When the sample size and the number of categories increase, the algorithm has a small cluster error too. Specific algorithm steps are as follows:

- (1) Initialize the first cluster center c_1 of samples O. Supposed $x_i \in O$ and $k=1$.

$$c_1 = \frac{1}{N} \sum_{i=1}^N x_i$$

- (2) $k = k + 1$, store the $k-1$ cluster centers of the last iteration c_1, \dots, c_{k-1} .
- (3) Get each sample as the k th cluster center, use K-means algorithm to optimize the initial cluster and get the best k types cluster, which center is c_1, \dots, c_k .
- (4) While $k = K$, stop computing; else go to step (2).

In order to optimize the structure of the PNN, we use Global K-means algorithm to select the number of hidden layers and their centers, the specific algorithm is as follows:

- (1) Set the number of cluster centers for each type samples. Initially $k = 1$.
- (2) According to the cluster Center k , use Global K-means algorithm to compute the cluster center vector of each type of samples.

- (3) (3)According to the cluster Center vector, design PNN classifier and calculate the classification error of the training samples.
- (4) (4)Test the classification error of training samples whether to achieve a given standard. Else $k = k + 1$, go to step (2).

The structure of PNN is showed as Fig. 2: the number of the first layer using radial basis neurons is the same with that of input sample vectors, which is average speed, road traveling time and change rate of instant speed; the second layer is the competitive level, equal to the number of training samples; the output layer corresponds to three traffic conditions which are smooth, stable and crowd.

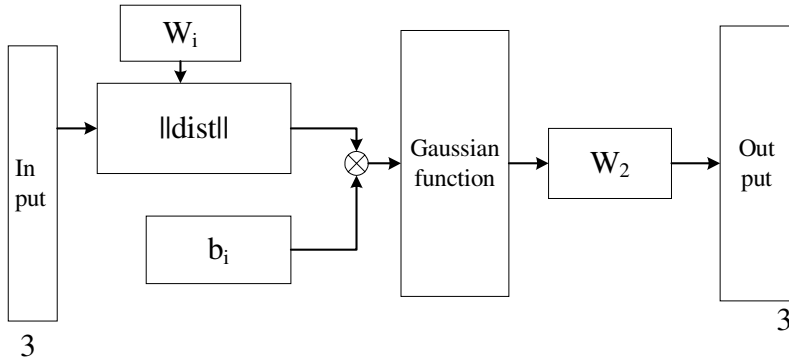


Fig. 2. The Structure of PNN

4.2 The Recognition of Traffic Condition Based on PNN and the Results

In order to study the accuracy of the method described, we have adopted two ways to test the traffic condition of different Guangzhou city’s roads, one is PNN, the other is practical manual testing. By comparing the actual traffic condition tested with the traffic condition of PNN based on floating car data, we can get the result of the algorithm.

According to floating car data of the Guangzhou City, select 504 sets of samples in a week of Tianhe road as the training sample data, including average speed, road traveling time and change rate of instant speed. There are 72 sets of samples in a day. Choose 280 sets as training samples(40 sets a day) and remaining 224 sets as testing samples(32 sets a day). Using Global K-means algorithm to optimize the number and position of PNN hidden layer center vector, determine the control parameters of kernel function in the light of experience. The optimized result is that the number of the first hidden layer center is reduced from 280 to 48, the compression rate is 82.9%. Traffic condition is divided into three states: smooth, stable and crowd.

Normalize the three dimensional sample vector as input; W_i weight is equal to the input vector transpose, the threshold $b_i = [-\log(0.5)]^{1/2} / spread^i$, $spread^i$ is the RBF expansion coefficient, after testing choose $spread^i$ as 0.7. The compared results of the two methods are as follows: the algorithm in this paper can identify 190 in the 224 sets of testing samples.the match degree is 84.8%, the degree that don’t match is 15.2%. Fig.3 is the compared result in the curve form of matching rates of testing samples, which can reflect the accuracy intuitively.

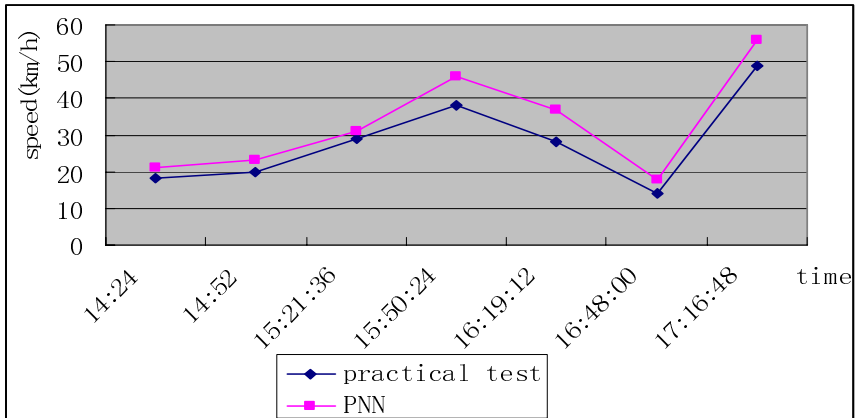


Fig. 3. Curve form of Matching Rate of Tianhe Road

5 Conclusions

After analyzing the Clustering algorithm based on PNN and Global K-means, this paper Optimized and designed the PNN classifier and applied it to recognize the urban roadway traffic condition combining the characteristic of traffic information collected by floating cars. The experiment results reveal that the method we present here can accurately reflect the the urban roadway traffic condition and has been used successfully in ITS Public Information Service Platform of Guangzhou city in China. Now, Guangzhou has developed the Taxi Management System and Vehicle Navigation System with the help of our method. The Taxi Management System realizes location and trace to 15000 taxis by Taxi GPS positioning techniques. The Vehicle Navigation System provides optimal driving paths to vehicles in virtue of the current traffic condition by analyzing and dealing with taxis GPS data.

References

1. Yu, L., Yu, L.: Traffic Incident Detection Algorithm for Urban Expressways Based on Probe Vehicle Data. *Journal of Transportation Systems Engineering and Information Technology* 8, 36–41 (2008)
2. Parkany, E., Xie, C.: A Complete Review of Incident Detection Algorithms & Their Development. R. In: Prepared for the New England Transportation Consortium, NETCR37 Project, pp. 0–7 (2005)
3. Zhang, C.B., Yang, X.G., Yan, X.P.: An Automatic Incident Detection Methodology for Freeway Using Floating Cars. *Journal of Wuhan University of Technology (Transportation Science & Engineering)* 30, 36–41 (2006)
4. Zhang, W.: Research on System Structure and key algorithm of ITS Common Information Platform in Guangzhou, South China University of Technology (2007)

5. Hofmann, M.O., Cost, T.L., Whylyry, M.: Model-based Diagnosis of the Space Shuttle Main Engine. *Artificial Intelligence for Engineering, Design, Analysis and Manufacturing* 6, 131–148 (1992)
6. Specht, D.F.: Probabilistic Neural Networks. *Neural Networks* 3, 109–118 (1990)
7. Likas, A., Vlassis, N., Verbeek, J.J.: The Gloabal K-means Clustering Algorithm. *Pattern Recognition* 36, 451–461 (2003)

Combined Neural Network Approach for Short-Term Urban Freeway Traffic Flow Prediction

Ruimin Li and Huapu Lu

Institute of Transportation Engineering,
Room 304, Heshanheng Building, Tsinghua University,
Beijing, 100084 China
lrmin@tsinghua.edu.cn

Abstract. Short-term traffic flow prediction is an essential component of urban traffic management and information systems. This paper presents a new short-term freeway traffic flow prediction model based on combined neural network approach. This model consists of two modules: a self-organizing feature map neural network and an Elman neural network. The former classifies the traffic conditions in a day into six patterns, and the later specifies the relationship between input and output to provide the prediction value. The inputs of the Elman neural network model include three kinds of data: the several time series of the prediction location, the historic data of the predictive time interval in the same weekdays, the time series of the adjacent location. The performances of the combined neural network model are validated using the real observation data from 3rd ring freeway in Beijing.

Keywords: Intelligent transportation system, Combined neural network, Short-term freeway traffic flow prediction, Spatiotemporal relationship.

1 Introduction

As the core of the urban Intelligent Transportation System (ITS), more and more traffic management centers have been established in China. On one hand, they collect traffic data through various traffic detection systems; on the other hand, they disseminate traffic information through variable message sign, broadcast, internet and other systems. At the same time, traffic information plays an important role in the successful advanced traffic management system (ATMS) and advanced traveler information systems (ATIS). Although the real-time traffic condition information can be accessed widely by the public via internet and other media sources, such information is less useful at the pre-trip planning or en-route stage because the traffic conditions are dynamically changing over time. Since drivers' decisions are influenced by expected future road network traffic conditions, it is clear that the most useful type of information for a driver to make a travel decision is reliable predictive traffic condition information. In order to improve the efficiency of ATMS and ATIS, the predictive information of traffic condition is necessary.

In the past three decades, substantial research has been done to develop the traffic prediction algorithms. For instance, the time series model [1,2], Kalman filtering model [3], nonparametric regression model [4,5], and neural network (NN) model [6,7] have been widely used to predict short-term traffic flow parameters. The neural network models used in the traffic prediction included back-propagation NN, RBF NN [8], and others [9,10]. Some comparison research between these models showed that no single prediction model may be accepted as the best one for real-time traffic prediction at all times.

Recently, several combined methods have emerged. Van Der Voort et al [11] utilized a hybrid method of the ARIMA model, which was used for predicting, and the Kohonen neural network, which was used for clustering. Park [12] presented a clustering-based RBF neural network model. Two clustering algorithms were compared, one is the Kohonen neural network, the other is the K-means method. Alecsandru et al [14] presented a hybrid traffic prediction system, in which the case-based reasoning was used to classify the traffic conditions and the ANN to predict the traffic condition value. It was found that the hybrid approach usually produced prediction value with smaller errors than those individual models.

Several researches indicate that a predictor has superior performance for a particular time period, so it is supposed that we will get improved performance by using different predictors in a particular period therefore arrive at more accurate prediction results in the whole period. The main task of this research is to find an appropriate methodology to predict the traffic flow parameters in the next time interval for a downstream location, based on the same observed traffic flow parameters in the former several intervals, in the same intervals of the former same weekdays as well as the former several intervals at the upstream locations. This research will verify whether the clustering of traffic condition in a day is useful for traffic prediction and prediction performance with spatiotemporal data outperforms that with only temporal data.

The presented research developed a combined neural network method for improving the performance of short-term traffic prediction systems under various recurrent traffic conditions. Firstly, we use a self-organizing feature map neural network to divide one day to different period with different traffic condition pattern, and secondly, an Elman neural network is used to specify the relationship between input and output to provide the predictive traffic information.

2 Combined Neural Network

2.1 Self-Organizing Feature Map Neural Network

SOFM NN. Self-organizing feature map neural network (SOFM), which is also called as Kohonen neural network, learn to classify input vectors according to how they are grouped in the input space. They differ from competitive layers in that neighboring neurons in the self-organizing map learn to recognize neighboring sections of the input space. Thus, self-organizing maps learn both the distribution (as do competitive layers) and topology of the input vectors they are trained on. So the SOFM NN may be used as a classifier to classify the traffic conditions in a day.

Input of SOFM NN. Because the traffic condition in urban freeway can be identified by the traffic flow volume, speed and occupancy, the input of SOFM neural network

includes the following variables, the traffic flow volume $Q(t)$, the average speed $V(t)$, and the average occupancy $O(t)$ in 1-hour increment. For a day, there are 24 records.

2.2 Elman Neural Network

Elman NN. The Elman neural network we consider here is a feed-forward network with three layers: an input layer, a hidden layer, and an output layer. The Elman NN differs from conventional feed-forward networks in that the input layer has a recurrent connection. Therefore, at each time step the output values of the hidden units are connected to the input ones. This process allows the network to memorize some information from the past, in such a way to have the advantageous time series prediction capability and to learn temporal patterns as well as spatial patterns.

Input of Elman neural network. A proper choice of the input dimension plays an important role in constructing an appropriate Elman neural network. As before-mentioned, in order to predict the traffic flow volume in the next interval for a downstream location, we consider the time of the day, the day of the week and the spatial relationship. Therefore, the input of the Elman NN includes the following three kinds of data.

Firstly, the traffic condition is periodically changed, for example, the traffic condition of Tuesday of this week is similar with the one of last week under normal traffic condition without special events, which is the inherent self-similarity of the traffic flow. So the past few weeks' data of the predicted interval are also included in the input of the Elman neural network. In the Elman neural network, the data of the predicted interval in the last 4 weeks were selected.

In addition, traffic flow data is a typical time series data, and the future traffic condition is influenced by the past traffic condition, as well as many existed predication methods used time-series traffic flow data to predict the future traffic condition. So the input of the network should include the time series data of the predicted location. The statistical autocorrelation function is used to determine the optimum size of past continuous discrete traffic flow volumes, eventually 4 past observations were used as the inputs.

Finally, in this paper, the Elman neural network accounting for the spatial characteristics of the freeway ring as well as the temporal evolution of traffic in the different location in the network. Generally speaking, the data from adjacent locations is useful to predict the data at the current location. Therefore, in order to predict more precisely, the information from adjacent locations should not be neglected. For example, in case of non-congestion the downstream measurement locations may depend on the upstream ones but in case of congestion the relation may be inverted. The number of the upstream locations is 2, as a rule of thumb, according to the selected freeway location.

In this paper, we will predict traffic flow volume in 5-min increment. The input variables of the Elman NN for traffic flow prediction include the flow volume $Q(t+5, k)_4$, $Q(t+5, k)_3$, $Q(t+5, k)_2$, $Q(t+5, k)_1$, $Q(t-15, k)$, $Q(t-10, k)$, $Q(t-5, k)$, $Q(t, k)$, $Q(t-5, k-2)$, $Q(t, k-2)$, $Q(t, k-1)$, while the output variable is $Q(t+5, k)$, with t representing the current time and k representing the prediction location in the network; while $k-1$, $k-2$ denotes the two nearer upstream location. The subscript 1,2,3,4 represent the same weekday in the last 4 weeks.

3 Case Study

After 2000, in order to improve freeway operation efficiency in Beijing, a traffic detection system has been implemented, covering three rings and eleven radial freeways. The system collects traffic data—including traffic volume, speed, and occupancy—from more than 400 microwave and ultrasonic detectors in 2-min increment.

The average traffic flow measurements along a three-lane freeway in the 3rd ring freeway of Beijing obtained from the Bureau of Beijing Traffic Management are used in this study to validate the combined neural network model. Three locations were selected, and each location has a microwave detector. For each location, the traffic data were collected from Sep 1, 2007, through Oct 31, 2007. For the purpose of this study, 5-min traffic flow measurements data were aggregated from the 2-min raw data, and only the data of each Tuesday were used in this study. The traffic conditions of the same weekday in different weeks are similar without holidays, so the data of Tuesday can demonstrate the effectiveness of the proposed model.

In order to incorporate the day of the week, a 5-min traffic flow data of each Tuesday are used in this study to demonstrate the effectiveness of the proposed model, resulting in 2,592 raw records. Each record includes 5-min traffic flow volume, 5-min traffic flow average speed, 5-min traffic flow average occupancy.

The 15-minute traffic flow pattern of the each Tuesday is shown in Fig.3, from which we will find that the pattern of Oct 2, 2007, a legal holiday day in China, is different with others, especially during 7:00-8:00, 12:30-13:30, 17:00-19:00 and 23:00-24:00. So traffic flow data on the holiday were excluded from the study.

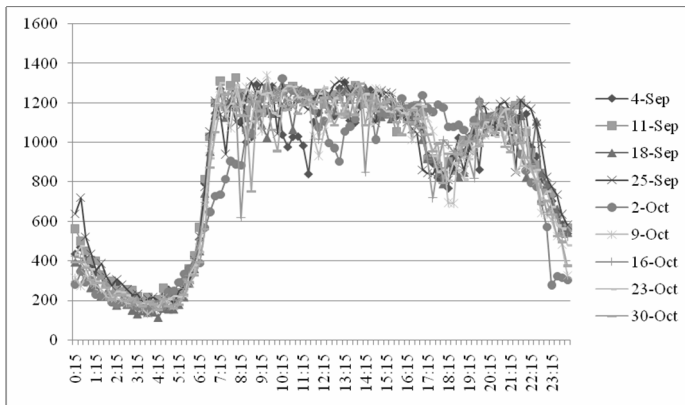


Fig. 1. 15-minute Traffic flow volume of 24 hours in nine Tuesdays

The main task of this numerical experiment is defined as predicting the traffic flow volume in the next time interval for the downstream location (Location 3), that is, $Q(t+5,3)$, based on the observed traffic flow data in the previous intervals, the last few weeks as well as from upstream locations (Locations 1 and 2).

The resulting traffic data used in this study were divided into training and test data sets for model development and testing. For the SOFM NN, 1 hour data were used to

cluster the traffic condition, and there are 8×24 records. For the Elman NN, the training data set included 852 records, and the testing set included 284 records. The training data sets were used to train the Elman neural network predictors for different periods. After the Elman neural network models were trained, the testing data set were used to test Elman NN performance. The prediction outputs were compared with the observed traffic flow data to test the performance of the Elman NN model.

4 Results and Discussion

Firstly, the SOFM NN was used to classify the traffic condition pattern in 1-hour increment. The traffic condition of each day was classified into six classes. The number of iteration of SOFM NN is 1000. The classification result, the time period of each traffic condition pattern and the number of training and testing data sets used in traffic prediction is shown in table. 1.

Table 1. Time period of each traffic condition pattern and the number of training and Testing data sets

Traffic pattern	1	2	3	4	5	6
Time period	7:00~17:00	17:00~20:00 21:00~22:00	23:00~24:00	20:00~21:00	6:00~7:00 22:00~23:00	0:00~6:00
Training data	360	144	36	36	72	204
Testing data	120	48	12	12	24	68

For comparison of the prediction performance of the models described, two indices, that is, mean absolute percentage error (MAPE), variance of absolute percentage error (VAPE) were selected and employed, as the MAPE and VAPE reflect the accuracy and stability of the predictor.

MAPE and VAPE are defined as follows,

$$MAPE = \frac{\sum_{t=0}^{N-1} (\frac{abs[V(t+1) - \hat{V}(t+1)]}{V(t+1)})}{N} \tag{1}$$

$$VAPE = \sqrt{\frac{N \sum_{t=0}^{N-1} (\frac{abs[V(t+1) - \hat{V}(t+1)]}{V(t+1)})^2 - [\sum_{t=0}^{N-1} (\frac{abs[V(t+1) - \hat{V}(t+1)]}{V(t+1)})]^2}{N(N-1)}} \tag{2}$$

where $V(t+1)$ =observed traffic flow volume for the time interval $t+1$, and $\hat{V}(t+1)$ = predicted traffic flow volume for the time interval $t+1$, N =number of intervals for prediction.

The prediction performance results are shown in Table 3, and the performance of the model was analyzed in detail.

Table 2. Performance Analysis of Prediction Results

Traffic pattern	1	2	3	4	5	6	Average ¹	Whole day ²
MAPE(%)-Elman	6.68	13.20	7.78	5.20	8.30	13.90	9.62	11.80
VAPE(%)-Elman	9.97	19.30	8.45	4.25	6.52	9.63	12.00	16.30

1—prediction performance with different Elman NN models for different traffic condition patterns.

2—with an Elman NN for the whole day.

From Table 2, it can be seen that in the six patterns, the individual Elman NN for each pattern has a visible better prediction performance than the Elman NN for the whole day under 4 condition patterns (1, 3, 4, 5) in terms of accuracy and stability, which is indicated by their MAPE and VAPE values. This indicates that after clustering the traffic condition, the applicability of individual Elman NN for each patter has been improved, that is, traffic condition classification is useful to improve the prediction accuracy and stability on these 4 patterns.

But under pattern 2 and 6, the individual Elman NN was worse than the Elman NN for the whole day, which may be as a result of the following reasons. For the pattern 2, from table.3 it was found that the classification results of different days have not been very consistent, which means that the traffic condition of pattern 2 in different days is not similar enough. For the pattern 6, from fig.1 it was found that the traffic condition of this pattern in each day has been also not similar enough.

Table 3. Classification results of traffic condition pattern 2 of eight Tuesdays

	Sep 4	Sep 11	Sep 18	Sep 25	Oct 9	Oct 16	Oct 23	Oct 30
17:00~18:00	2	5	2	5	2	5	2	2
18:00~19:00	2	2	2	5	5	2	2	2
19:00~20:00	2	2	2	4	2	2	2	2
21:00~22:00	1	4	2	4	2	4	2	2

Finally, for the total 284 prediction data, the MAPE and the VAPE of integrated individual Elman NN are 9.62%, and 12.00% respectively, and an Elman NN for the whole day are 11.80% and 16.30%. It was also found that for the whole day, the individual Elman NN has a better prediction performance than the Elman NN for the whole day. With such accuracy, the combined model could be considered as a potential model for field implementation.

In order to verify the influence of different input variables to the prediction performance of the Elman NN, three scenarios different kinds of input variables were also tested, including following:

- (a)—inputs of Elman NN includes $Q(t+5,k)_4$, $Q(t+5,k)_3$, $Q(t+5,k)_2$, $Q(t+5,k)_1$, $Q(t-15,k)$, $Q(t-10,k)$, $Q(t-5,k)$, $Q(t,k)$, $Q(t-5,k-2)$, $Q(t,k-2)$, $Q(t,k-1)$

- (b)—inputs of Elman NN includes $Q(t+5,k)_4$, $Q(t+5,k)_3$, $Q(t+5,k)_2$, $Q(t+5,k)_1$, $Q(t-15,k)$, $Q(t-10,k)$, $Q(t-5,k)$, $Q(t,k)$,
- (c)—inputs of Elman NN includes $Q(t-15,k)$, $Q(t-10,k)$, $Q(t-5,k)$, $Q(t,k)$.

Table 4. Prediction error comparisons of Elman NN model with different input variables

Traffic pattern	1	2	3	4	5	6	Average ¹	Whole day ²
MAPE(%)-Elman(a)	6.68	13.20	7.78	5.20	8.30	13.90	9.62	11.80
VAPE(%)-Elman(a)	9.97	19.30	8.45	4.25	6.52	9.63	12.00	16.30
MAPE(%)-Elman(b)	7.32	15.17	12.1	5.11	9.01	14.15	10.53	11.0
VAPE(%)-Elman(b)	13.7	18.86	13.2	4.07	10.24	10.58	13.66	16.83
MAPE(%)-Elman(c)	7.37	10.56	4.87	5.72	9.28	14.64	9.64	10.63
VAPE(%)-Elman(c)	13.23	21.06	5.53	3.89	6.01	12.83	14.18	15.17

1—prediction performance with different Elman NN models for different traffic condition patterns.

2—with an Elman NN for the whole day.

From the table.4, it can be seen that whatever the input variables were, the MAPE of different Elman NN models for different traffic patterns, which is shown in column 8, are smaller than that of an Elman NN model for the whole day, which is shown in column 9, also the VAPE. Comparison results indicate that after traffic condition pattern division, the accuracy and stability of the prediction can be improved.

From the column 8 in table.4, it can be seen that the (a) scenario has best prediction accuracy and stability, and the (c) scenario is only little worse than (a) scenario in prediction accuracy, but the prediction stability is not as good as (a) scenario. The prediction accuracy of (b) scenario is worse than (c) scenario, but the stability is better than that of (c) scenario. It is shown that with the decrease of input variables, the stability of prediction became worse.

Among these six traffic condition patterns, the prediction accuracy of (a) scenario is best under 1,5,6 patterns, the prediction accuracy of (b) scenario is best under 4 patterns and the prediction accuracy of (c) scenario is best under 2,3 patterns. From which it can be seen that there is no one prediction model suiting for every traffic condition pattern. So we may conclude that improved prediction performance may be achieved in the whole period by using different prediction models in different periods.

With the data from the same system, Liguang Sun et al [15] studied the performance of a Combined Short Term Traffic Flow Forecast Model which consists of discrete Fourier transform model (DFT), autoregressive model (AR) and neighborhood regression model (NR). For traffic flow volume forecasting in Tuesday, the MAPE of the whole day is 12.3%, worse than 9.62%, the MAPE of the proposed model in this paper. Also the MAPE of the commuting time, 9.3% is worse than 8.28%.

5 Conclusions

In this study, a combined neural network prediction model of a SOFM neural network and an Elman neural network was presented for short-term freeway traffic flow volume prediction and evaluated with freeway traffic data from the 3rd ring freeway of

Beijing. It was found that after clustering the traffic conditions into several classes in a day by SOFM NN, the individual Elman NN prediction models for each traffic pattern outperformed only one Elman NN for the whole day. On the other hand, the proposed model incorporates both the time of the day, the day of the week of the prediction time and the spatial relationship of the predicted location, which may improve the prediction performance in some time period in a day.

In this research, an arbitrary number of both the past weeks and the related location of the predicted location respectively were used. More research efforts should be undertaken on how to selection such parameters to improve the prediction performance. The models proposed here were only tested with normal traffic volume cases. In future, whether the proposed model is suitable for the abnormal traffic condition or not should be studied.

Acknowledgment. This work was supported by Committee of Beijing Science and Technology (Grant No. D07020601400704) and National High-Tech Research and Development Program of China (863 Program)(Grant No.2007AA11Z233).

References

1. Ahmed, S.A., Cook, A.R.: Analysis of Freeway Traffic Time-Series Data Using Box-Jenkins Techniques. *Transp. Res. Rec.* 722, 1–9 (1979)
2. Williams, B.M., Durvasula, P.K., Brown, D.E.: Urban Freeway Traffic Flow Prediction: Application of Seasonal Autoregressive Integrated Moving Average and Exponential Smoothing Models. *Transp. Res. Rec.* 1644, 132–141 (1998)
3. Okutani, I., Stephanedes, Y.J.: Dynamic Prediction of Traffic Volume Through Kalman Filtering Theory. *Transp. Res. B.* 18, 1–11 (1984)
4. Smith, B.L., Demetsky, M.J.: Multiple-Interval Freeway Traffic Flow Forecasting. *Transp. Res. Rec.* 1554, 136–141 (1996)
5. Davis, G.A., Nihan, N.L.: Nonparametric Regression and Short-Term Freeway Traffic Forecasting. *J. Transp. Eng.* 117, 178–188 (1991)
6. Dougherty, M.S., Kirby, H.R., Boyle, R.D.: The Use of Neural Networks to Recognize and Predict Traffic Congestion. *Traffic Eng. & Control* 34, 311–314 (1993)
7. Smith, B.L., Demetsky, M.J.: Short-term Traffic Flow Prediction: Neural Network Approach. *Transp. Res. Rec.* 1453, 98–104 (1994)
8. Park, B., Carroll, J.M., Urbank II, T.: Short-term Freeway Traffic Volume Forecasting Using Radial Basis Function Neural Network. *Transp. Res. Rec.* 1651, 39–46 (1998)
9. Jiang, X.M., Adeli, H.: Dynamic Wavelet Neural Network Model for Traffic Flow Forecasting. *J. Transp. Eng.* 131, 771–779 (2005)
10. Zheng, W.Z., Lee, D.H., Shi, Q.X.: Short-term Freeway Traffic Flow Prediction: Bayesian Combined Neural Network Approach. *J. Transp. Eng.* 132, 114–121 (2006)
11. Voort, M.V., Dougherty, M., Watson, S.: Combining Kohonen Maps with ARIMA Time-Series Models to Forecast Traffic Flow. *Transp. Res. C.* 4, 307–318 (1996)
12. Park, B.: Clustering-Based RBF Neural Network Model for Short-Term Freeway Traffic Volume Forecasting. In: Hendrickson, C.T., Ritchie, S.G. (eds.) *Applications of Advanced Technologies in Transportation Engineering*, School of Civil Engineering, Purdue University, West Lafayette, Indiana (1988)

13. Alecsandru, C., Ishak, S.: Hybrid Model-Based Model and Memory-Based Traffic Prediction System. *Transp. Res. Rec.* 1879, 59–70 (2004)
14. The MathWorks, Inc. Matlab r2007a
15. Sun, L.G., Dong, S., Chang, T.H., Lu, H.P.: Combined Short Term Traffic Flow Forecast Model for the Beijing Traffic Forecast System of Olympic 2008. *Transp. Res. Rec.* (accepted)

Facial Expression Recognition in Video Sequences

Shenchuan Tai and Hungfu Huang

Department of Electrical Engineering, National Cheng Kung University,
Tainan, Taiwan

hhf93d@ddcmc.ee.ncku.edu.tw

Abstract. This paper proposes a system for the facial expression recognition. Firstly, we perform noise reduction by a median filter of facial expression image. Then, a cross-correlation of optical flow and mathematical models from the facial points are used. To define these facial points of interest in the first frame of an input face sequence image, which utilize manually marker. The facial points were automatically tracked by a cross-correlation, which is based on optical flow, and then extracted the feature vectors. The mathematical model extracts features from the feature vectors. An ELMAN neural network was applied to classify expressions. The performances of the proposed facial expressions recognition were computed by Cohn–Kanade facial expressions database. This proposed approach achieved a high recognition rate.

Keywords: Noise reduction, Median filter, Optical flow, ELMAN, Neural network.

1 Introduction

The face is a unique and important feature of human beings, conveying identity and emotion. When we look at a person's face, we not only discern who it is but also process other information, such as the expression, gender, ethnicity, and age. A successful expression classification method has many potential applications such as human identification, human–computer interface, computer vision approaches for monitoring people, passive demographic data collection, etc.

The median filter introduced by Tukey [1] in the 1970s, has been used extensively for image noise reduction and smoothing. Median filters are especially good at removing impulsive noise from images. The particular nonlinearity of the median filter permits it to smooth an image without the degree of blurring that a linear filter with similar smoothing characteristics can introduce.

The face can send many subtle signals. For example, an array of facial expressions—a smile of happiness, a frown of sadness or disapproval, wide-open eyes of surprise, or lips curled in disgust—all show a wide range of emotions. Research by

Ekman et al. [3] have identified six basic emotions that people can identify from facial expressions with high accuracy.

The Facial Action Coding System (FACS) [4,5] is currently the most widely used method in recognizing facial expressions. FACS encodes the contraction of each

facial muscle (stand alone as well as in combination with other muscles) that changes the appearance of the face, and it has been used widely for the measurements of shown emotions.

Gizatdinova and Surakka [6] used feature-based method for detecting landmarks from facial images. The method was based on extracting oriented edges and constructing edge maps at two resolution levels. Edge regions with characteristic edge pattern formed landmark candidates.

An optical-flow based approach [7,8] is sensitive to subtle changes in facial expression. Action unit (AU) combinations in the brow and mouth regions were selected for analysis if they occurred at a minimum twenty-five times in the database.

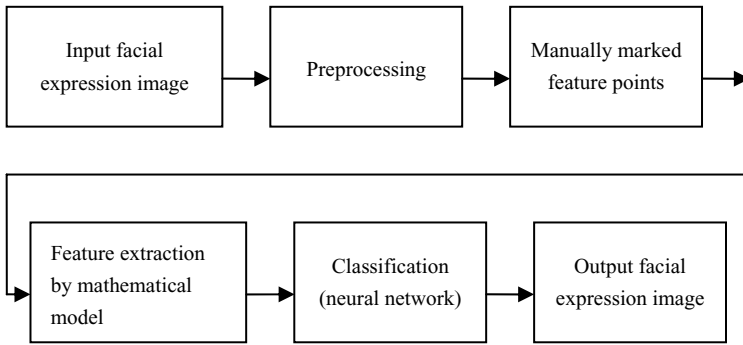


Fig. 1. The facial expression recognition system

Selected facial features were automatically tracked by using a hierarchical algorithm for estimating optical flow. Image sequences were randomly divided into training and test sets.

We performed a preprocessing of facial expression image. Then, proposed seven feature mathematical models extracted from seventeen facial feature points, and formed a feature vector for each expression. These features were used to train an ELMAN neural network classifier to classify input feature vectors into one of the six basic emotions. This approach flow shows in Fig. 1.

The paper is organized as follows. Section 2 is the preprocessing of facial expression image. Section 3 finds seventeen facial points. We extracted seven features from seventeen facial points by optical flow and mathematical models in section 4. While in section 5 these features were used to train and test an ELMAN neural network. Finally, the results are described in section 6.

2 Preprocessing

2.1 Noise Reduction

In image processing, the median filter [1,2] has been proved to be a very effective method in removing noise. When encountering an image corrupted with noise you will want to improve its appearance for a specific application.

2.2 Median Filter Algorithm

For an image I of size $M \times N$ and an observation X of the size same as I . An U is the impulsive noise with random values.

$$X = I + U \quad (1)$$

A median filter replaces a pixel by the median of all pixels in the neighbourhood:

$$y[m, n] = \text{median}\{X[m, n], (m, n) \in W\} \quad (2)$$

where $m = 1 \dots M$, $n = 1 \dots N$, and W is a predetermined window. Generally, W is chosen to be 3×3 , 5×5 , or 7×7 .

3 Facial Feature Point Marker

In the first frame, 17 feature points were manually marked with a computer-mouse around facial landmarks (Fig. 2).



Fig. 2. Seventeen facial points

4 Feature Extraction

4.1 Facial Points Tracking

Since the database of this approach used image sequences, tracking facial points were decomposed temporally into successive two-frame matching problems. In addition, we could track the facial points that were faster than repeating the whole detection process. The facial points tracking used cross-correlation of optical flow [9-13]. Each point was the center of a 13 by 13 flow window that included horizontal and vertical flows. A cross-correlation based on optical flow method is used to automatically tracking the facial points in the sequence image.

Cross-correlation of a 13 by 13 window in the front frame, and a 23 by 23 window at the next frame were calculated and positioned with maximum cross-correlation of two windows, were estimated as position of feature point at the next frame. Each feature point was calculated by subtracting its standard position in the first frame from its current standard position. All the feature points' positions were patterned by position of the top of nose.

4.2 Feature Extraction by Mathematical Model

The feature extraction used mathematical model description. Since the paper using the Cohn–Kanade database consists of expression sequences of subjects, starting from a neutral expression and ending in the peak of the facial expression. Hence, seven features were extracted from facial point position in the first frame and the end frame. The feature mathematical models are as below:

- Size of Eye:

$$((Y5 - Y6) + (Y9 - Y10)) / 2 \tag{3}$$

- Width of Eye:

$$((X4 - X3) + (X8 - X7)) / 2 \tag{4}$$

- Eyebrow to Iris Distance:

$$((Y1 - Y11) + (Y2 - Y12)) / 2 \tag{5}$$

- Width of Mouth:

$$X15 - X14 \tag{6}$$

- Size of Mouth:

$$Y16 - Y17 \tag{7}$$

- Philtrum:

$$Y13 - Y16 \tag{8}$$

- Eye to Cheek Distance:

$$((Y11 - Y13) + (Y12 - Y13)) / 2 \tag{9}$$

where X_n is the position in Figure 1 of feature point n [$n \in 1 \dots 17$] in x -axis and Y_n is the position in Figure 1 of feature point n [$n \in 1 \dots 17$] in y -axis. X and Y represent x -axis and y -axis respectively in the two-dimension area.

5 The ELMAN Neural Network Model

All features were classified as one of the six basic emotions by means of the ELMAN neural network system [14-19]. It belongs to recurrent networks which differentiate other networks, and it is the additional connection from the hidden unit to itself.

The recurrent connection allows detecting and generating time-varying patterns. The simplest method of recurrent network is ELMAN network.

We used two hidden layers, and the number of input layer units must be equal to the number of extracted features and the output layers correspond to six kinds of facial expressions. The highest value will be indicated to the corresponding facial expression. The network architecture is shown as Fig. 3.

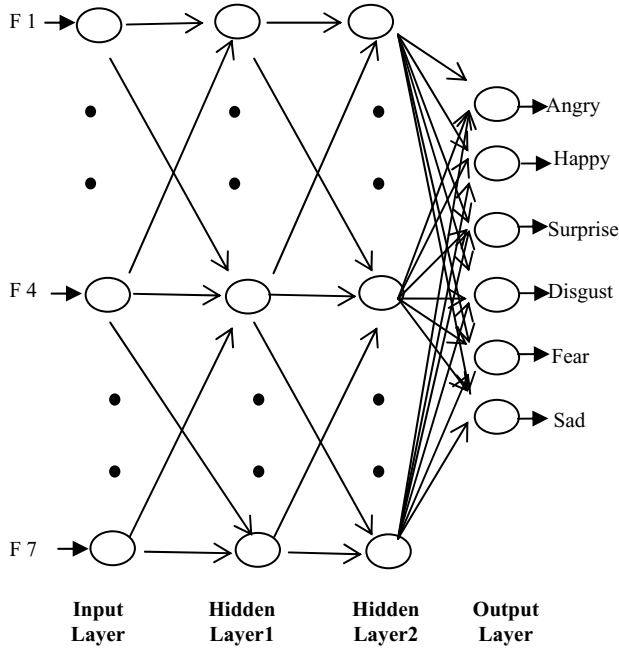


Fig. 3. The ELMAN neural network architecture

6 Results

6.1 Data Set

The Cohn-Kanade facial expression database consists of 100 university students aged from 18 to 30, of which 65% was female, 15% were African-American, and 3% were Asian or Latino. Subjects were instructed to perform a series of 23 facial displays, in which six were based on descriptions of prototypic emotions (i.e., anger, disgust, fear, happy, sad, and surprise). Image sequences from neutral to target displays were digitized into 640x490 pixel arrays.

6.2 Simulation Result

To show the efficiency of our method, extensive experiments were done on the Cohn-Kanade facial expression database to test, and to train. However, not all of the six facial expressions sequences were available to us in all subjects, we used only a

subset of fifty-five subjects, for which at least four of the sequences were available. Every facial expression of the subject had no regular number of image sequences. All subjects must have the faces in the same illumination and on the same scale. The choice of training and test set obeyed the following rules: (1) The training set selects forty-five subjects randomly; (2) The test set selects ten subjects randomly; (3) The test was repeated five times; (4) Different testing and training subjects were changed each time. Table 1 shows the average recognition rate.

Table 1. Average recognition rate

		Accuracy (%)	
		Before preprocessing	After preprocessing
Recognition Feature	Angry	90.06	90.74
	Happy	100	100
	Surprise	96.66	96.9
	Sad	89.42	90.62
	Disgust	96.7	97.22
	Fear	84	86.78
Average (%)		92.8	93.71

6.3 Result Comparison

Six methods for recognizing expressions of novel individuals were compared in Table 2. All these methods were tested based on the Cohn-Canade database and uses a similar way to divide the database. In [20], a method a Nearest-Neighbor algorithm, which achieved an average recognition rate of 89.13%. In [21], a method consists of three modules. First, face detection were used. Second, Gabor wavelet and AdaBoost were applied to select feature that presents a face image. Finally, Gabor features selected by AdaBoost were fed into NKFDA to classify, with a recognition rate of 85.6%. In [22], that used Relevance Vector Machines (RVM) as a novel classification technique for the recognition of facial expressions and an accuracy of 90.84% was achieved.

Table 2. Comparison result of six methods

Other methods	Accuracy (%)
J.M. Buenaposada et al. [20]	89.13
Huchuan et al. [21]	85.6
Datcu et al. [22]	90.84
S.C. Tai et al. [23]	92.0
Seyedarabi et al. [9]	91.6
H.C. Lu et al. [24]	92.26
Our Method	93.71

In [23], an approach based on the optical flow and mathematical model achieved an average recognition rate of 92.0%. In [9], a facial expression recognition system, which was based on the facial features extracted from facial characteristic points in frontal image sequences, was developed. This method achieved an average recognition rate of 91.6%. In [24], that used pixel-pattern-base texture feature and SVM. This method achieved an average recognition rate of 92.26%. In our method, we have obtained a higher recognition rate, 93.71%.

7 Conclusion

This paper presented a high-performance facial expression recognition method. There were first preformed by the median filter which reduced noise and by the logarithm image processing which enhanced image quality. Then, we proposed a method for the estimation of the facial features, which used manually marker facial points. Moreover, the cross-correlation of optical flow does facial point tracking and mathematical models do feature extraction from the facial points. The ELMAN neural network trained and tested seven features, the experiments showed that our method performs better than others on the Cohn-Canade database. This system achieved more efficient facial expression recognition.

Acknowledgment. The authors would like to thank the Cohn-Kanade Technical Agent. We are also grateful to the anonymous reviewers for their valuable comments.

References

1. Tukey, J.W.: Nonlinear (nonsuperposable) Methods for Smoothing Data. In: Proc. Congr. Rec. EASCOM, vol. 74, p. 673 (1974)
2. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 2nd edn. Prentice-Hall, Englewood Cliffs (2002)
3. Ekman, P., Friesen, W.V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., Lecompte, W.A., Pitcairn, T., Ricci-Bitt, P.E., et al.: Universals and Cultural Differences in the Judgments of Facial Expressions of Emotion. *J. Pers. Soc. Psychol.* 53, 712–717 (1987)
4. Ekman, P., Friesen, W.V.: Facial Action Coding System. Consulting Psychologist Press, Palo Alto (1978)
5. Ekman, P., Friesen, W.: Pictures of Facial Affect. Consulting Psychol. (1976)
6. Gizatdinova, Y., Surakka, V.: Feature-Based Detection of Facial Landmarks from Neutral and Expressive Facial Images. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 28, 135–139 (2006)
7. Cohn, J.F., Zlochow, A.J., Lien, J.J., Kanade, T.: Feature-Point Tracking by Optical Flow Discriminates Subtle Difference in Facial Expression. In: Proc. IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, pp. 396–401 (1998)
8. Mase, K.: Recognition of Facial Expression from Optical Flow. *IEICE Transactions, E* 74, 3474–3483 (1991)
9. Seyedarabi, H., Aghagolzadeh, A., Khanmohammadi, S.: Recognition of Six Basic Facial Expressions by Feature-Points Tracking using RBF Neural Networks and Fuzzy Inference

- System. In: The IEEE International Conference on Multimedia & Expo (ICME 2004), Taipei, Taiwan (June 2004)
10. Seyedarabi, H., Aghagolzadeh, A., Khanmohammadi, S.: Facial Expressions Recognition from Static Images Using Neural Networks and Fuzzy Logic. In: The 2nd Iranian Conference on Machine Vision and Image Processing (MVIP 2003), Tehran, vol. 1, pp. 7–12 (2003)
 11. Black, M.J., Yacoob, Y.: Recognizing Facial Expressions under Rigid and Non-Rigid Facial Motions. In: International Workshop on Automatic Face and Gesture Recognition, Zurich, pp. 12–17 (1995)
 12. Rosenblum, M., Yacoob, Y., Davis, L.S.: Human Emotion Recognition from Motion Using a Radial Basis Function Network Architecture. In: Proceedings of the Workshop on Motion of Non-rigid and Articulated Objects, Austin, TX (1994)
 13. Medioni, G., Kang, S.B.: *Emerging Topics in Computer Vision*. Prentice-Hall, Englewood Cliffs (2004)
 14. Christodoulou, C., Georg, M.: *Applications of Neural Networks in Electromagnetics*. Artech House (2001)
 15. Medsker, L., Jain, L.C.: *Recurrent Neural Networks Design and Applications*. CRC, Boca Raton (1999)
 16. Samarasinghe, S.: *Neural Networks for Applied Sciences and Engineering: From Fundamentals to Complex Pattern Recognition*. AUERBACH (2006)
 17. Kumar, V.V., Krishnamurthy, A.K., Ahalt, S.C.: Phonetic-to-Acoustic and Acoustic-to-Phonetic Mapping Using Recurrent Neural Networks. *Applications of Artificial Neural Networks (SPIE)* 1469, 484–494 (1991)
 18. Topouzelis, K., Karathanassi, V., Pavlakis, P., Rokos, D.: Dark Formation Detection Using Recurrent Neural Networks and SAR Data. In: *Proc. of SPIE*, vol. 6365, pp. 111–117 (2006)
 19. Lee, S.W., Lee, E.J.: Integrated Segmentation and Recognition of Connected Handwritten Characters with Recurrent Neural Network. In: *Proc. SPIE*, vol. 2660, pp. 251–261 (1996)
 20. Buenaposada, J.M., Munoz, E., Baumela, L.: Recognising Facial Expressions in Video Sequences. *Pattern Anal. Applic.* 11, 101–116 (2008)
 21. Lu, H., Wang, Z., Liu, X.: Facial Expression Recognition Using NKFDA Method With Gabor Features. In: *Proceedings of the 6th World Congress on Intelligent Control and Automation*, Dalian, China (2006)
 22. Datcu, D., Rothkrantz, L.J.M.: Facial Expression Recognition with Relevance Vector Machines. In: *IEEE International Conference on Multimedia and Expo*, pp. 193–196 (2005)
 23. Tai, S.C., Huang, H.F., Chung, K.C.: Automatic Facial Expression Discrimination System. *Far East Journal of Electronics and Communication* 1, 23–31 (2007)
 24. Lu, H.C., Huang, Y.J., Chen, Y.W., Yang, D.I.: Real-time Facial Expression Recognition Based on Pixel-pattern-based Texture Feature. *Electronics Letters* 43(17) (August 16, 2007)

An AFSA-TSGM Based Wavelet Neural Network for Power Load Forecasting

Dongxiao Niu¹, Zhihong Gu¹, and Yunyun Zhang²

¹ School of Business Administration, North China Electric Power University,
Beijing 102206, China

² College of Economics Management, North China Electric Power University,
Baoding 071003, China

{Zhihong Gu, laobing618}@126.com

Abstract. An intelligent methodology for power load forecasting was developed. In this forecasting system, wavelet neural network techniques were used in combination with a new evolutionary learning algorithm. The new evolutionary learning algorithm introduced the Tabu Search Algorithm and Genetic Mutation Operator into Artificial Fish Swarm Algorithm (AFSA) to construct a hybrid optimizing algorithm, and is thus called ASFA-TSGM. The hybrid algorithm can greatly improve the ability of searching the global excellent result and the convergence property and accuracy. The effectiveness of the ASFA-TSGM based WNN was demonstrated through the power load forecasting. The simulated results show its feasibility and validity.

Keywords: Wavelet neural networks, Tabu search, Genetic mutation operator, AFSA, Power load forecasting.

1 Introduction

As the power market develops gradually, power companies attach more and more importance to accurate load forecast. There are many kinds of normal methods for load forecast. Among them the artificial neural network (ANN) [1-3] has been widely applied to power load forecast for its good nonlinear mapping ability. The multilayer perception (MLP) [4], along with the back-propagation (BP) training algorithm, is probably the most frequently used type of neural network in practical applications. Unfortunately, these ANNs have some inherent defects, such as low learning speed, existence of local minima, and difficulty in choosing the proper size of network to suit a given problem. To solve these defects, we combine wavelet theory with it and form a wavelet neural network (WNN) whose activation functions are drawn from a family of wavelets. The WNN has shown surprising effectiveness in solving the conventional problem of poor convergence or even divergence encountered in other kinds of neural networks. It can dramatically increase convergence speed [5-7].

How to determine the structure and parameters of the neural networks promptly and efficiently has been a difficult issue all the time in the field of neural networks research [9]. This paper tries to apply the ASFA-TSGM that introduced the Tabu Search Algorithm and Genetic Mutation Operator into Artificial Fish Swarm

Algorithm for training the WNN. Artificial Fish Swarm Algorithm (AFSA) [8-9] is a new kind of intelligence optimization algorithm, which is inspired by the behavior of fish. This algorithm has some benefits such as robustness against local extreme, fast convergence and very flexible in practice. But as the complexity and scope of optimization problem expanding, it is difficult to get satisfied result by single optimization algorithm [10-11], so as to AFSA. AFSA has several disadvantages such as the blindness of searching at the later stage and the poor ability to keep the balance of searching, which reduce its probability of searching the best result. To overcome these problems, this paper introduced the Tabu Search Algorithm and Genetic Mutation Operator into AFSA to construct a hybrid optimizing algorithm. A two-point tabu search operator is proposed to avoid the iteration during optimization, and it is adopted to simulate the activities of a single fish such as pursuing the historically optimal fish, pursuing the current optimal fish and converging to the center of the other fishes. The non-uniformity mutation operator used in genetic algorithm is used to represent the individual activity of searching food. The whole fish swarm communicate and cooperate by the above four activities, thereby generating the hybrid AFSA based on tabu search and genetic mutation operator. The hybrid algorithm can adjust the searching range adaptively and have better ability to keep the balance of searching, and can avoid circuitous searching. And then we use it training WNN for power load forecasting. The experimental results show that the proposed algorithm is significantly superior to normal BP algorithm and original AFSA.

The paper is organized as follows. The WNN for power load forecasting is described in Section 2. In section 3, the basic AFSA and the improved algorithms are explained. In section 4, designs of WNN by ASFA-TSGM are simulated and applied to power load forecasting. Finally, section 5 concludes the paper.

2 Wavelet Neural Network for Power Load Forecasting

The WNN employed in this study is designed as a three-layer structure with an input layer, wavelet layer (hidden layer) and output layer. The topological structure of the WNN is illustrated in Figure 1.

The activation functions of the wavelet nodes in the wavelet layer are derived from a mother wavelet $\psi(x)$. Suppose that the Fourier transform of the square-integrable

function $\psi(x) \in L^2(R)$ is $\hat{\Psi}(w)$, and it satisfies the condition: $\int_{-\infty}^{+\infty} |w|^{-1} \left| \hat{\Psi}(w) \right|^2 dw < +\infty$ [12], so function $\psi(x) \in L^2(R)$ is the mother wavelet. Order

$$\Psi_{ab}(t) = \Psi((t-b)/a) / \sqrt{|a|} \tag{1}$$

so $\Psi_{ab}(t)$ is the wavelet base which is generated by mother wavelet and depended on parameter a and b , in the formula, a is the translation factor, b is the expansion factor.

Suppose the nonlinear time series transform function is $f(t) \in L^2(R)$, definite the wavelet transform as:

$$\Psi_{ab}(t) = \int_{-\infty}^{+\infty} f(t)\Psi((t-b)/a)dt / \sqrt{|a|} \tag{2}$$

Choose wavelet base series finite linear combination to approximate time series function[13-14],

$$g(t) = \sum_{i=1}^N w_i \Psi((t-b_i)/a_i) \tag{3}$$

In the formula, $g(t)$ is the predicted value of dynamic error. As in Fig.1, w_1, w_2, \dots, w_N are weight coefficients. θ is combination of all parameter, t is the input channel, $g(t)$ is the output channel, the mean square error function of predicted value is set up as the object function $C(\theta)$, so there is:

$$C(\theta) = \frac{1}{2} \sum_{i=1}^N [g(t) - f(t)]^2 \tag{4}$$

Order $\xi_i = (t-b_i)/a_i$, $\xi(t) = g(t) - f(t)$, so the learning procedure is as follows:

- 1) choose suitable wavelet $\Psi(t) = \cos(1.75t)e^{-t^2/2}$;
- 2) initialize parameter θ_k ;
- 3) compute gradient of $C(\theta)$;

$$h(w_i) = \partial C / \partial w_i = \sum_{k=1}^N e_k \cos(1.75\xi_i) e^{-\xi_i^2/2}$$

$$h(b_i) = -\sum_{k=1}^N e_k w_i a_i^{-1} \{1.75 \sin(1.75\xi_i) + \xi_i \cos(1.75\xi_i)\} e^{-\xi_i^2/2}$$

$$h(a_i) = -\sum_{k=1}^N e_k w_i \xi_i a_i^{-1} \{1.75 \sin(1.75\xi_i) + \xi_i \cos(1.75\xi_i)\} e^{-\xi_i^2/2}$$

4) use recurrence to seek optimum. It can use step-by-step checking method to confirm n . If the error of fitting is less than D , then n begins from 1 and compute ξ_1 . If $\xi_1 < D$, then $n=1$, or increase n to 2; If $\xi_2 < D$, then $n=2$, or continue, and let $\xi_n^* < D$ when $n = n^*$, so the optimal value of schema n can be confirmed and written as n^* .

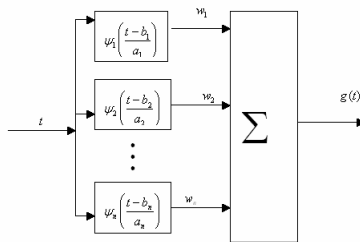


Fig. 1. Structure of WNN

3 AFSA Based on Tabu Search and Genetic Mutation Operator

AFSA performs a parallel search in the solution space to find the optimum solution by simulating the behavior of fish including searching food, congregate and follow. In other words, this algorithm searches the optimum solution based on the cooperation and competition of the fish individuals. The key of designing AFSA lies on simulating behaviors of a single fish, which needs to define the appreciable visual distance VISUAL, the move step length STEP and the crowd factor δ . As space is limited, the basic theory of AFSA will not be shown here which was explained in articles [8] and [9].

This paper proposed an improved AFSA based on tabu search and genetic mutation operator. It adopts the two-point tabu search operator to simulate such behaviors of a single fish as pursuing the historically optimal fish, pursuing the current optimal fish and converging to the center of the other fishes, and adopts the non-uniformity genetic mutation operator to simulate the behavior of searching for food. The concrete theory of the proposed method will be discussed as follow.

3.1 Introduction of Tabu Search

Tabu Search is a kind of stochastic sub-elicitation searching algorithm proposed by Glover [15], which searches the solution space through its memory ability and the rule of expectation. The basic theory of tabu search is: Supposed that there is a solution and an adjacent field, it will find an optimal solution in this adjacent field and write it as *ans*, then make the present optimal solution $ans^* = ans$ and search the local optimal solution ans' in the adjacent field nearing this present optimal solution. But this optimal solution may equal to the former one, in order to avoid this kind of circulation, the tabu search algorithm set a tabu table to remember the latest operation. If the present operation is in the table, it will be forbidden, otherwise *ans* will be replaced by ans' . At this moment, the object function of *ans* may be better than that of ans' , so the tabu search algorithm can accept bad solution. But if those especially beneficial operations improved the present optimal solution, they could be released from the tabu table by the rule of expectation to find better solution fast.

To the optimizing problem of consecutive variables, the prerequisite of tabu search is that the solution space should be discrete (determining the number of adjacent fields *Nadj*). So this paper adopts the strategy of adjacent fields tabu, it means that if a point was remembered and stored in the tabu table, then all the points in the adjacent field whose center is that point would be forbidden. After that the solution space would become discrete, and the tabu search algorithm could be used just given the length of the tabu table *Ntabu*. This paper constructed a two-point tabu search operator to search the solution space without repetition.

3.2 Two-Point Tabu Search Operator

Supposed that there is a searching field, and its lower limit is $L = (l_1, l_2, \dots, l_n)$ and the upper limit is $U = (u_1, u_2, \dots, u_n)$. Any point A (a single fish) of the optimizing swarm (fish swarm) could be seen as a slip surface, and its position could be set as

$X = (x_c, y_c, x_h) = (z_1, z_2, z_3)$. The other optimizing point B could be either the optimal point of the present swarm or the historical optimal point of swarms or the geometrical center of the present swarm, and its position could be set as $O = (o_1, o_2, o_3)$. $\alpha_{\min}, \alpha_{\max}$ are separately the lower limit and the upper limit of the optimizing parameter α , and they can be determined by L and U .

After the scale of the optimizing parameter α is determined, the process of searching a new point X_{new} using points A and B can be defined as follow:

$$X_{new} = O + \alpha(O - X) \tag{5}$$

Firstly, make $\alpha = \alpha_{\max}$, we can get a new point according to formula (5). If this new point is feasible, then determine which adjacent field it belongs to and judge whether this adjacent field k is in the tabu table. If this adjacent field k is not in the tabu table, it demonstrates that the optimizing operator makes a successful searching and put this adjacent field k in the tabu table, and if the tabu table is full, we can update it according to the first-in first-out principle; If this adjacent field k is in the tabu table, then judge whether the rule of expectation is met, and if it is, the searching is successful, otherwise make $\alpha = 0.5\alpha$ and return to formula (5) to get a new point again, and repeat the above judgment; If this new point is unfeasible, then make $\alpha = 0.5\alpha$ and return to formula (5) to get a new point again, and repeat the above judgment until reaching a successful searching. If the searching is still not successful when α becomes a very small number (10^{-3}), then make $\alpha = \alpha_{\min}$ and repeat the above steps until successful; if it is still not successful when α becomes much smaller, then the optimizing operator fail. Figure 2 shows searching process of the two-point tabu search operator. From figure 2 we can see that the new points generated by the operator lay in zone I when $\alpha \in [0, \alpha_{\max}]$, and when $\alpha \in [\alpha_{\min}, 0]$ they lay in zone II.

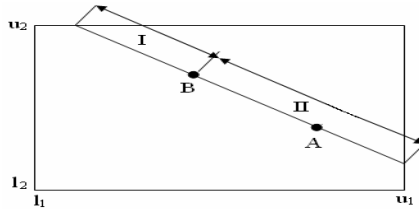


Fig. 2. Two-point tabu search operator

3.3 Behaviors of a Single Fish

1) Pursuing the historically optimal fish. The basic AFSA sets up a call-board to record the information of the historically optimal fish, but it is not used for searching for the global optimal solution. In order to use the information of the historically optimal fish, this paper simulated the behavior of pursuing the historically optimal fish. Supposed that the present position of a single fish is $X = (z_1, z_2, z_3)$ and that

of the historically optimal fish is $G = (g_1, g_2, g_3)$, we can take the present single fish as the point A of the two-point tabu search operator, and take the historically optimal fish as the point B, then the two-point tabu search operator can be used to do the searching process, if it is successful, the present fish should be replaced with the new point, otherwise the present fish just searches for food. L_5 in Figure 3 shows the process of the fifth fish pursuing the historically optimal fish.

- 2) Pursuing the current optimal fish. Supposed that the present position of a single fish is $X = (z_1, z_2, z_3)$ and that of the optimal fish in the present fish swarm is $LOC = (loc_1, loc_2, loc_3)$, we can take the present single fish as the point A of the two-point tabu search operator, and take the optimal fish of the present fish swarm as the point B, then the two-point tabu search operator can be used to do the searching process, if it is successful, the present fish should be replaced with the new point, otherwise the present fish just searches for food. L_6 in Figure 3 shows the process of the sixth fish pursuing the optimal fish of the present fish swarm.

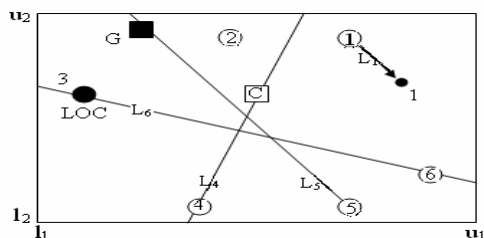


Fig. 3. Individual activities of two-dimensional fish swarm

- 3) Converging to the center of the other fishes. Supposed that the present position of a single fish is $X = (z_1, z_2, z_3)$ and the center position of the other fishes is $C = (c_1, c_2, c_3)$, we can take the present single fish as the point A of the two-point tabu search operator, and take the center position C as the point B, then the two-point tabu search operator can be used to do the searching process, if it is successful, the present fish should be replaced with the new point, otherwise the present fish just searches for food. L_4 in Figure 3 shows the process of the fourth fish moving to the geometrical center.
- 4) Searching for food using non-uniformity genetic mutation operator. This paper uses the non-uniformity genetic mutation operator [16] to construct the behavior of searching for food. This operator has only one parameter b , and in the earlier stage of iteration, it can assure the fish swarm to search for food in broader range, and in the later stage of iteration, it can assure the fish swarm to convergence around the optimal solution, which can assure reaching higher searching precise. The computing formula is as follow:

$$z'_k = \begin{cases} z_k + \Delta(t, u_k - z_k) \\ z_k - \Delta(t, z_k - l_k) \end{cases} \quad (6)$$

Where $\Delta(t, y) = y \cdot \text{rand}(1 - t/T_{\max})^b$, t is the present evolution time and T_{\max} is the maximum evolution time, rand is random numbers among (0, 1), b is a system parameter which determines the dependence degree of stochastic disturbance to the evolution time t , usually $b = 2$. Supposed that the present position of a single fish is $X = (z_1, z_2, z_3)$, it can execute once mutation and finish once searching for food according to formula (6), and if the new position does not meet the constraint conditions, then searching for food again until new feasible point is found. Taking the training of WNN as example, the concrete iteration steps of the hybrid AFSA-TSGM algorithm is explained as follow.

3.4 Steps of AFSA-TSGM Algorithm

- 1) Set the dimension of limit vector $n = 3$, the number of adjacent fields $N_{adj} = 100$ and make the solution space as discrete fields, set the length of tabu table $N_{tabu} = 20$. Set the number of fishes $N = 6$ and generate 6 fishes meeting the constraint conditions, and put those adjacent fields which the 6 fishes belonging to in the tabu table. Initialize the iteration time $t = 0$ and give the maximum iteration time T_{\max} .
- 2) Compute the object function of the 6 fishes $S_i (i = 1, 2, \dots, 6)$ according to formula (4), compute the optimal fish LOC of the present fish swarm and historically optimal fish G , and record the information of the historically optimal fish on the call-board.
- 3) Single fishes update their positions by pursuing the historically optimal fish, pursuing the current optimal fish, converging to the center of the other fishes and searching for food, and compute the historically optimal fish G and update the information on the call-board.
- 4) Make $t = t + 1$, if $t \geq T_{\max}$, then end the algorithm, otherwise return to step 3).

4 Experiment

4.1 Process of Samples

The method of this paper was applied to forecast loads of a power system, and it was compared with the original AFSA. The programming system was Matlab 7.0. Data of the above power system from Aug.2002 to Jul.2004 were chosen as training samples, and that from Aug.2004 to Dec.2004 were chosen as test samples.

In order to keep neurons from saturation phenomenon and improve the forecasting precise, data of loads should be normalized as follow:

$$x'_t = (x_t - x_{\min}) / (x_{\max} - x_{\min}) \quad (7)$$

The output should be reverted as follow:

$$x_t = x'_t (x_{\max} - x_{\min}) + x_{\min} \quad (8)$$

Table 1. Comparison of the 24-hour forecasting error (%)

Time of forecasting	AFSA-TSGM	AFSA
00:00:00	0.61	0.47
01:00:00	0.38	0.48
02:00:00	0.43	0.59
03:00:00	1.61	1.94
04:00:00	0.26	0.53
05:00:00	0.54	0.74
06:00:00	0.31	1.46
07:00:00	1.28	1.72
08:00:00	0.87	1.15
09:00:00	1.73	2.10
10:00:00	1.89	2.38
11:00:00	1.47	2.17
12:00:00	1.79	2.66
13:00:00	2.24	3.18
14:00:00	1.52	2.57
15:00:00	2.08	3.02
16:00:00	1.65	2.46
17:00:00	0.46	1.25
18:00:00	1.62	0.79
19:00:00	1.70	2.73
20:00:00	2.76	3.92
21:00:00	1.68	3.04
22:00:00	4.25	5.85
23:00:00	4.24	6.07

where x_{\max} and x_{\min} are separately the maximum load and the minimum load of samples. The other influencing factors could be processed as the method proposed by article [16].

4.2 Choice of Parameters

Some parameters of the algorithms in this paper were set as follow: To the normal BP algorithm, the training time was initialized as 100, and the learning probability was set as 0.2, and the initial attenuation probability was set as 5×10^{-4} , and the learning speed was set as 0.01. To the original AFSA, two groups of limit vectors were approximately given as: $U_1=(90,80,45)$, $L_1=(10,10,25)$, $U_2=(50,60,45)$, $L_2=(10,0,25)$; to every group of limit vectors, generate 10 groups of initial fish swarm (the number of fishes in every fish swarm is 6) randomly; the appreciable visual distance $VISUAL=25$, the move step length $STEP=50$ and the crowd factor $\delta=6$; and the maximum iteration time $T_{\max}=200$. To the AFSA-TSGM algorithm, adopt the 20

initial fish swarms generated as above, the maximum iteration time $T_{\max}=200$; the number of adjacent fields $N_{adj}=100$, and the length of tabu table $N_{tabu}=20$; the system parameter $b=2$.

4.3 Error Test and Result Analyses

This paper adopts the average relative error e as the criterion judging the forecasting precise:

$$e = \frac{1}{n} \sum_{i=1}^n \left| \frac{A(i) - F(i)}{A(i)} \right| \times 100\% \quad (9)$$

where $A(i)$ and $F(i)$ are separately the real load and the forecasted load.

In order to test the capability of the AFSA-TSGM algorithm, it was compared with the original AFSA. Table 1 and Figure 4 showed the comparing results. Table I showed the forecasting errors of the three algorithms to the day of Aug 23, 2004, and it could be concluded that the average of relative errors of AFSA-TSGM was 1.56%, which was smaller than 2.22% of AFSA. Figure 4 showed the forecasting curve and the real load curve, and it also showed that AFSA-TSGM had achieved better forecasting precise.

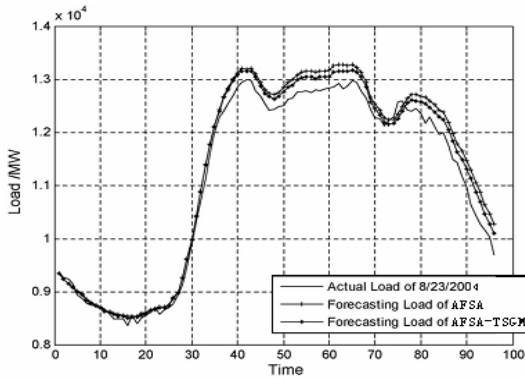


Fig. 4. Curves of forecasted and real load

5 Conclusion

A wavelet neural network approach based on AFSA-TSGM is developed for power load forecasting. The approach uses the wavelet theory to construct a wavelet neural network to increase the convergence speed, and takes the novel kind of AFSA-TSGM optimization algorithm to train wavelet neural network. The feasibility and effectiveness of this new approach is validated and illustrated by the study case of modeling power load forecasting.

Acknowledgement. This work is supported by National Natural Science Foundation of China (NSFC) (70671039) and by Program for New Century Excellent Talents in University (NCET-07-0281).

References

1. Wang, Z.Y., Guo, C., Cao, Y.J.: A Method for Short Term Load Forecasting Integrating Fuzzy-rough Set with Artificial Neural Network. *Proceedings of the CSEE* 25, 7–11 (2005)
2. Xie, H., Cheng, H.Z., Zhang, G.L., et al.: Applying Rough Set Theory to Establish Artificial Neural Networks for Short Term Load Forecasting. *Proceedings of the CSEE* 23, 1–4 (2003)
3. Niu, D.X., Chen, Z.Y., Xing, M., Xie, H.: Combined Optimum Gray Neural Network Model of the Seasonal Power Load Forecasting with the Double Trends. *Proceedings of the CSEE* 22, 29–32 (2002)
4. Chen, D.S., Jain, R.C.: A Robust Backpropagation Learning Algorithm for Function Approximation. *IEEE Transaction on Neural Networks* 5, 467–479 (1994)
5. Pati, Y.C., Krishnaprasad, P.S.: Analysis and Synthesis of Feedforward Neural Networks Using Affine Wavelet. *IEEE Transaction on Neural Networks* 4, 73–75 (1993)
6. Zhang, J., Walter, G.G., Miao, Y., Lee, W.N.W.: Wavelet Neural Networks for Function Learning. *IEEE Transaction on Signal Process* 4, 1485–1497 (1995)
7. Delyon, B., Juditsky, A., Benveniste, A.: Accuracy Analysis for Wavelet Approximations. *IEEE Transaction on Neural Networks* 6, 332–348 (1995)
8. Li, X.L., Qian, J.X.: Studies on Artificial Fish Swarm Algorithm based on Decomposition and Coordination Techniques. *Journal of Circuits and Systems* 8, 1–6 (2003)
9. Li, X.L., Shao, Z.J., Qian, J.X.: An Optimizing Method based on Autonomous Animals: Fish Swarm Algorithm. *Practice and Theory for System Engineering* 11, 32–38 (2002)
10. Xia, W.J., Wu, Z.M.: An Effective Hybrid Optimization Approach for Multi-objective Flexible Job-shop Scheduling Problem. *Computers & Industrial Engineering* 48, 409–425 (2005)
11. Salhi, S., Queen, N.M.: A Hybrid Algorithm for Identifying Global and Local Minima when Optimizing Functions with Many Minima. *European Journal of Operational Research* 155, 51–67 (2004)
12. Daubechies, I.: The Wavelet Transform, Time-frequency Localization, and Signal Analysis. *IEEE Transaction on Inform. Theory* 36, 961–1000 (1990)
13. Chui, C.K.: *Wavelet: A Tutorial in Theory and Applications*. Academic Press, Boston (1992)
14. Ryotaro, K.: Minimum Entropy Methods in Neural Network: Competition and Selective Responses by Entropy Minimization. *IEEE International Joint Conference of Neural Network* 5, 219–225 (1993)
15. Ji, M.J., Tang, H.W.: Global Optimizations and Tabu Search based on Memory. *Applied Mathematics and Computation* 159, 449–457 (2004)
16. Niu, D.X., Xing, M., Meng, M.: Research on ANN Power Load Forecasting Based on United Data Mining Technology. *Transactions of China Electro technical Society* 19, 62–68 (2004)

Comparative Analyses of Computational Intelligence Models for Load Forecasting: A Case Study in the Brazilian Amazon Power Suppliers

Liviane P. Rego, Ádamo L. de Santana, Guilherme Conde, Marcelino S. da Silva,
Carlos R. L. Francês, and Cláudio A. Rocha

Laboratory of High Performance Networks Planning,
Federal University of Pará-R. Augusto Correa,
Belém 66075-110, Brazil
{lrego, adamo, conde, marcelino, rfrances}@ufpa.br,
alex@bcc.unama.br

Abstract. One of the most desired aspects for power suppliers is the acquisition/sale of energy for a future demand. However, power consumption forecast is characterized not only by the variable of the power system itself, but also related to socio-economic and climatic factors. Hence, it is imperative for the power suppliers to design and correlate these parameters. This paper presents a study of power load forecast for power suppliers, comparing application of techniques of wavelets, time series analysis methods and neural networks, considering long term forecasts; thus defining the future power consumption of a given region. The results obtained proved to be much more effective when compared to those projected by the power suppliers based on specialist information, thus contributing to the decision making for acquisition/sale of energy at a future demand.

Keywords: Power load forecast, Time series analysis, Wavelets, Neural networks.

1 Introduction

Power load forecasting has always been the essential part of an efficient power system planning and operation [1,2]. With it, power suppliers can satisfactorily estimate the purchase of energy based on the future demand and prices, minimizing the difference between the amounts of energy bought and consumed.

With respect to load forecasting, due to corporative reasons, power suppliers are, at times, unable to retrieve or apply exogenous variables to the model, due to the fact that these variables are hard and/or expensive to obtain. Use of some attributes are also discarded for predictive analysis, for not enabling their estimation accurately and thus, not being able to deal with the forecasting with the algorithms and techniques used, especially as the period of forecasting increases. We will consider in the study presented here the alternative for load forecasting when no data, other than the energy consumption itself, is available for the power suppliers.

The contributions of this work follow the study and improvement of the load forecasting estimations, using statistical and computational intelligence models, specifically regression, neural networks and wavelets; inducing, in the course, new variables and approaches. Thus providing with a comparative analysis of their effectiveness and applicability.

This paper is organized as follows: the regression methods for load forecast are subject of section 2. Section 3 presents the neural networks model used and the definition of its parameters. Section 4 presents the wavelets model applied for the forecasting. In section 5, final remarks of the paper are presented.

2 Time Series and Regression Model for Load Forecasting

The data available for the analysis with regression model correspond to the total power consumption. The study used the historical power consumption data available for the period from January 1991 to December 2006.

As discussed in previous studies [3], the consumption time series is tendentious and non stationary. The series, by studying its correlograms, does not achieve stationarity even on successive differentiations. Thus, a new approach that consists of partitioning the series in twelve annual series corresponding to the months from January to December was used (Fig. 1). These now partitioned series when studied presented stationarity.

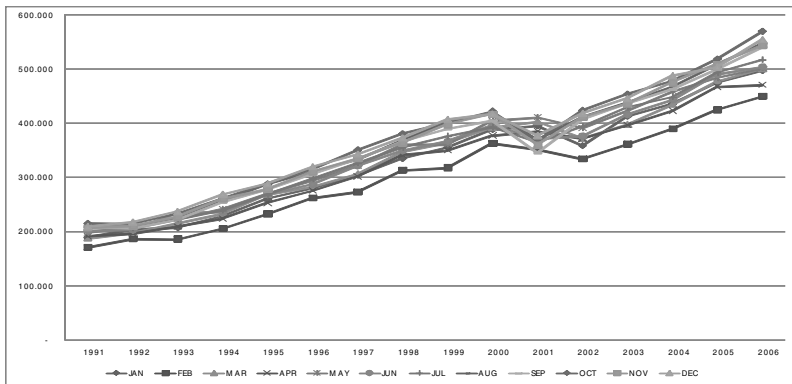


Fig. 1. Historical data of the power consumption from 1991 to 2006 separated monthly

An almost linear growth could then be observed in the series throughout time, apart from the period ranging from 2001 to 2002, characterized by the occurrence of a national measure for energy rationing; which drastically reduced the power consumption [4].

The estimator for the consumption prediction uses a multiple regression analysis (see [5] and [6]), based on the value of the consumption at a previous time and two

additional terms. The general formula of the multiple regression model can be specified as (1):

$$Y_i = A_0 + A_1 X_{1i} + A_2 X_{2i} + \dots + A_k X_{ki} + u_i \quad (1)$$

where, Y is a column vector, with dimension $n \times 1$; X is a matrix of size $n \times k$, that is, with n observations and k variables; with the first column representing the intercept A_0 ; A is a vector with $k \times 1$ unknown parameters; u is a vector with $n \times 1$ disturbances.

Although the non stationarity issue is solved by partitioning the data into annual series, the introduction of this approach might bring a loss of knowledge from events exogenous (e.g. effects originated from loss or acquisition of contracts by the energy suppliers, projects or government managements, etc.) to the standard behavior of the system, occurred during the eleven months apart from the next instance of the analyzed month.

To account for the impact of such events and thus obtain a more adjustable value for the prediction, a variable, obtained from a factorial analysis (for a more complete view on factorial analysis see [7]), was added to quantify the annual trend of the consumption according to its behavior and to condense the information and trends occurred in the year.

The factorial analysis was made over the twelve annual series; retrieving from the analysis a single factor which best represents the series (around 99.6%) and, consequently, the annual behavior.

The second term added acts with respect to the containment of the impact from anomalies in the historical data of the power consumption (from June 2001 to February 2002) by adding a new binary artificial variable to the monthly series; indicating the presence or absence of a historical value influenced by this occurrence, assigning values 1 or 0, respectively.

Not only the period when the rationing measure was installed is treated here, but also the months that had followed that period, which persisted with a decrease in the power consumption, until the series returned to its normality.

The performance of the model will be evaluated according to mean absolute percentage error (MAPE), calculated according to (2).

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left(\frac{|y_i - \hat{y}_i|}{y_i} \right) \times 100\% \quad (2)$$

where N is the number of existing samples, y is the real historical value and \hat{y} is the estimated value. The results from the neural networks and the wavelets methods, presented in sections 3 and 4, respectively, are also based on MAPE.

An initial test was made using the data from 1991 to 2004 and then estimating the consumption values for the years of 2005 and 2006 (Fig. 2), that presented an error of approximately 1.76%, a value inferior to all of the statistical methods used by the national power suppliers, which runs around 4.1%.

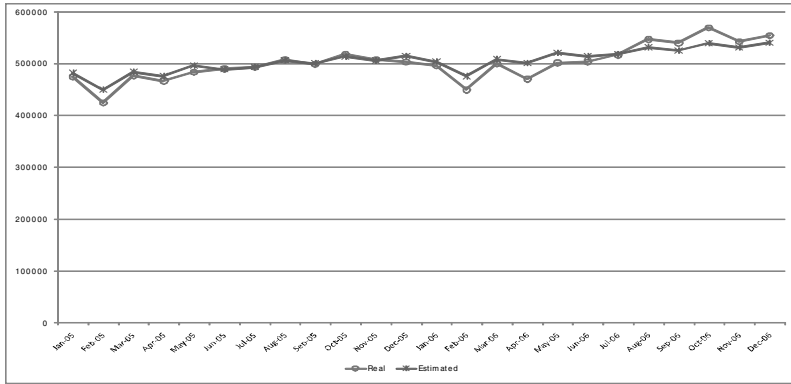


Fig. 2. Real and estimated values of the power consumption from Jan/05 to Dec/06

Once the effectiveness of the regression estimation model for the data series is verified, a projection of its behavior was made for the years 2007 and 2008 (Fig. 3).

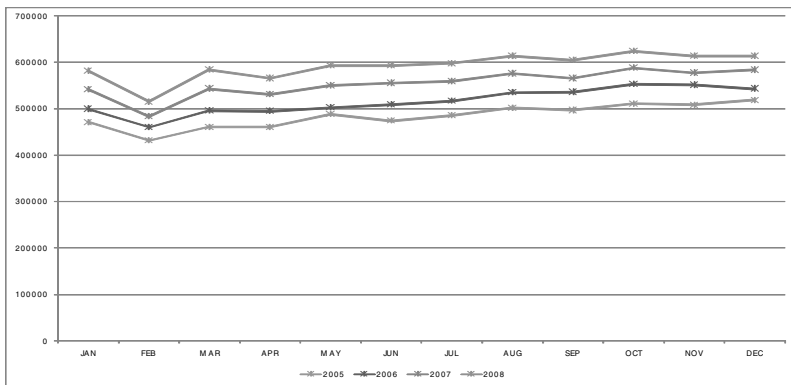


Fig. 3. Estimated values of the power consumption for the years 2005 to 2008

The results achieved by the implemented neural networks model, as well as a comparative analysis of the results obtained by both the techniques will be shown next.

3 Artificial Neural Networks Model

The artificial neural networks used in this work carry out the forecasting of the power consumption, whose input variables are its historical data decomposed in twelve series, as previously described in section 2, and the variables of *date* and the *growth rate* from consecutive years were also used. The ANN was fed with the value of the consumption and its three previous values, using a “windowing” technique with size three. Fig. 4 illustrates the inputs and outputs selected for modeling the ANN used.

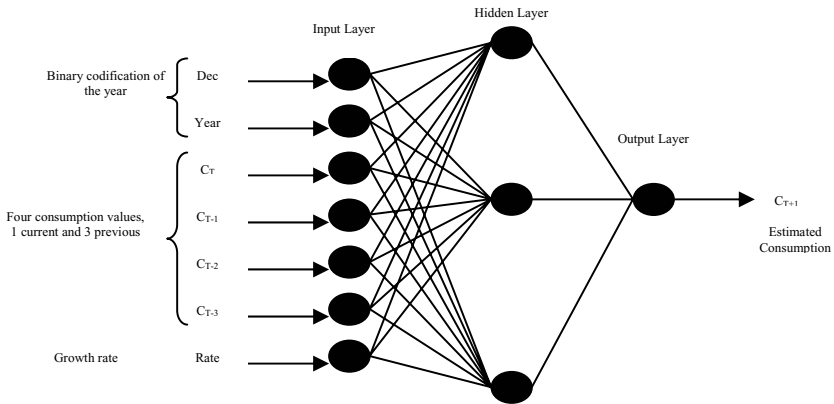


Fig. 4. RNA architecture used

The ANN architecture chosen was the feedforward multilayer perceptron network (MLP), with one hidden layer, due to its wide use on predictive systems [8,9] and the training algorithm chosen for the learning of the MLP network was the backpropagation [10] and Levenberg-Marquardt [11], with modifications for the dynamic adjustment of the configuration parameters. After preliminary experiments, the learning rate in Levenberg-Marquardt algorithm was fixed at 0.02.

The simulations were made for each of the twelve series, identifying and selecting the best MLP network for each. The simulations were made with two data sets.

First, the MLP networks used as trainings sets the historical data from 1991 to 2004, and as test data from year 2005. After the training process, the forecast of the consumption values for the years 2005 and 2006 were made with the backpropagation and Levenberg-Marquardt algorithms (Fig. 5 and 6, respectively).

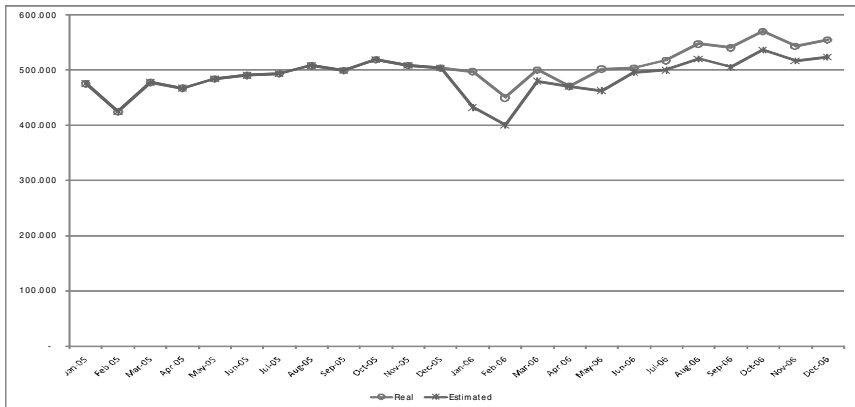


Fig. 5. Real and estimated values of the power consumption from Jan/05 to Dec/06 obtained by the MLP with backpropagation

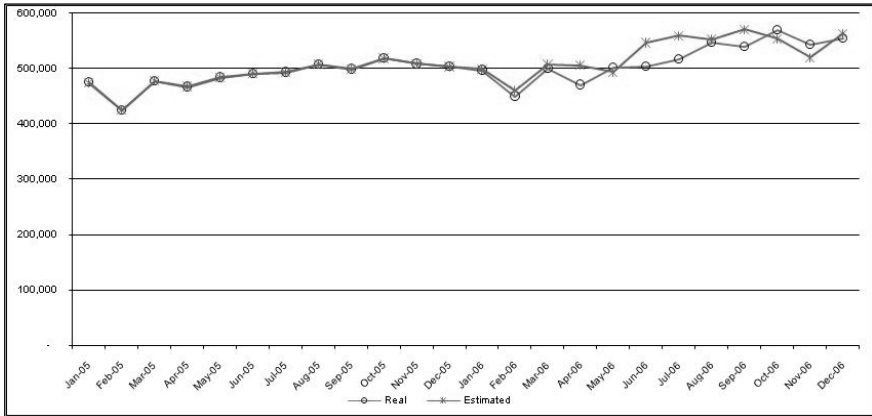


Fig. 6. Real and estimated values of the power consumption from Jan/05 to Dec/06 obtained by the MLP with Levenberg-Marquardt

The obtained results for the forecasts presented residual errors in both the cases, of approximately $2 \times 10^{-4}\%$ with backpropagation and $26 \times 10^{-3}\%$ with Levenberg-Marquardt, for the year 2005. However, for 2006 it generated an error of 6.06% with backpropagation and 3.75% with Levenberg-Marquardt.

The neural network models were applied again; having now as training set the values of the historical data from 1991 to 2005 and, for test, the data of 2006. The forecast of the consumption values was then made with both backpropagation and Levenberg-Marquardt for the years 2006, 2007 and 2008 (Fig. 7 and 8, respectively).

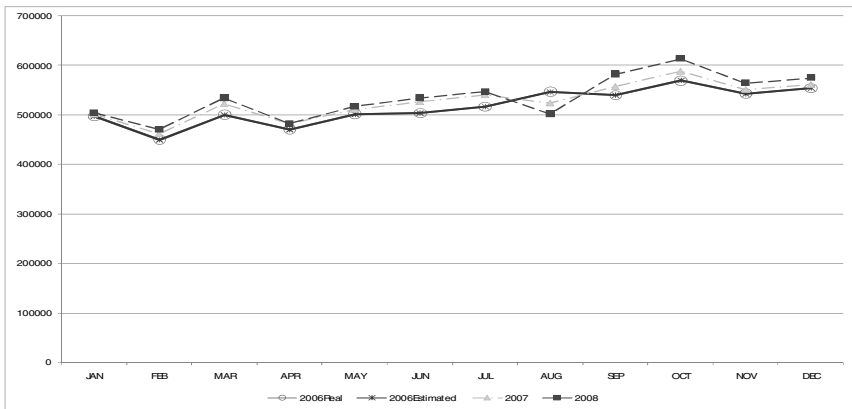


Fig. 7. Estimated values of the power consumption for the years from 2006 to 2008 obtained by the MLP with backpropagation

The forecast of the consumption for 2006, shown in Fig. 7 and 8, also presented residual errors, of approximately $1 \times 10^{-4}\%$ with backpropagation and $13 \times 10^{-4}\%$ with Levenberg-Marquardt.

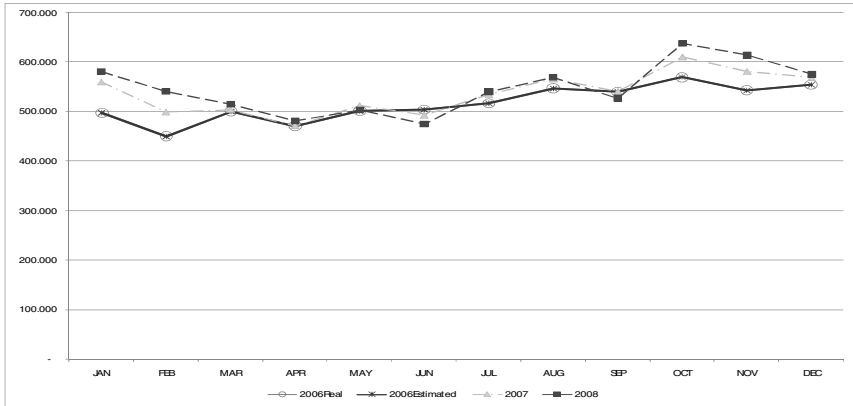


Fig. 8. Estimated values of the power consumption for the years from 2006 to 2008 obtained by the MLP with Levenberg-Marquardt

The results presented here showed that the MLP networks present an exceptional result when studying a short-range forecast, usually periods from six months to one year. However, when studying longer periods of time, the reliability of the values decreases drastically, presenting anomalous values (Fig. 5, 6, 7 and 8) of consumption after the first year of forecast; such values do not agree with those from the knowledge of specialists in the domain. For long-range forecasts, the model based on regression techniques, presented in section 2, yielded better results, producing series with a good behavior (Fig. 3); and also a good adjustment, although inferior to the one obtained by the MLP model.

4 Wavelets Forecasting Model

The idea when applying wavelets on time series analysis and forecasting is to decompose the original time series signal into smoother components and then to apply the most appropriate prediction method for each component, individually. In this context, the high frequency components are best to explain near future trends, while the low frequency components contains the general tendencies of the series and can be used to tell the long term trend [12].

In the model studied here, a non-decimated wavelet transform (NWT) was applied, using the Daubechies [13] function of order 5. The original data series was initially decomposed into sets of approximation and details; furthermore, in the tests made, we found that a decomposition of level two was the best for the problem studied (Fig. 9).

The components were then studied as individual time series. Given that the time series multiple regression method obtained the best overall results (considering a joint application for both short and long term forecasting), when compared to the neural networks model, the former was chosen to be applied in the wavelet time series analysis, instead of using a wavelet network [14] approach for it.

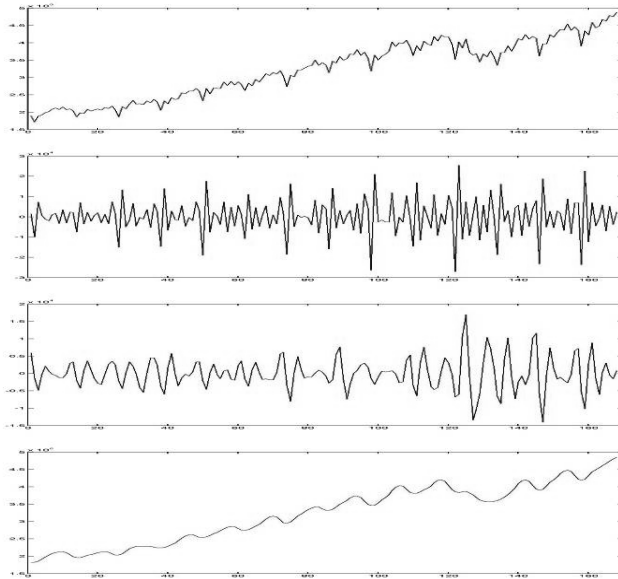


Fig. 9. Original series (top) and the wavelet components at level two of decomposition

A multiple regression analysis similar to the one presented in section 2 was then applied for each data series (approximation and details) separately. The predictive results obtained for each component were then added to build the final results of the model. An overview of the forecasting system is presented in Fig. 10.

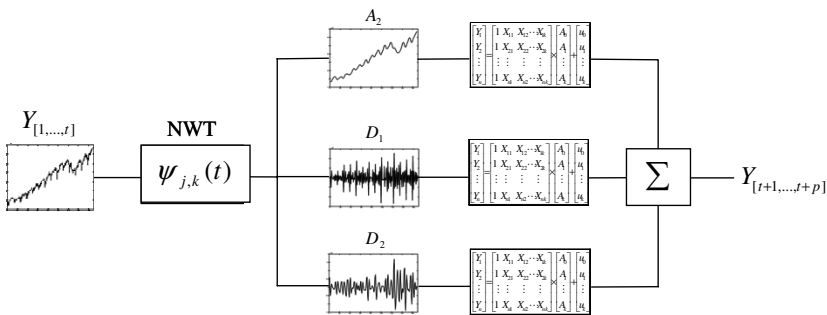


Fig. 10. Wavelet/multiple regression forecasting system

The model obtained a total forecasting error of 0.72%, improving the results obtained by the multiple regression method, while maintaining the good behavior that the technique provides for long term forecasting. The values obtained (real and estimated) are shown in Fig. 11.

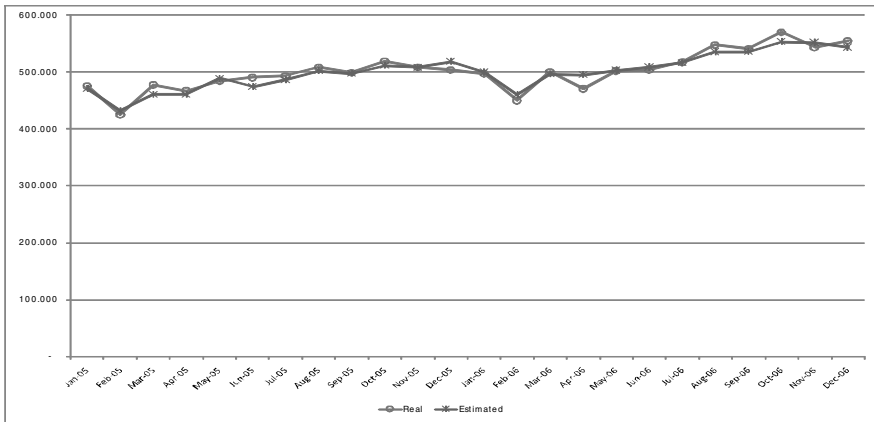


Fig. 11. Real and estimated values of the consumption from Jan/05 to Dec/06 obtained by the wavelet model

5 Final Remarks

The objective of this work is to study forecasting models for the power consumption. In this paper, three techniques for power load forecasting were applied: by using time series methods, neural networks and wavelets models.

In the tests, it was observed for all the estimators studied a good capacity of adjustment and prediction, presenting percentage errors below the ones currently seen by the traditional methods used by the power suppliers; with the wavelets model improving the multiple regression method, outperforming its results. This improvement in the error reduction represents, evidently, a considerable economy for energy purchase in a future market.

As a distinguishing aspect, it is also pointed out that, as it could be observed by the obtained results, the estimator based on neural models presented an exceptional performance for a short range forecast of up to one year, presenting a residual error value; but, when considering a forecasting for longer periods, the models produce anomalous values with respect to the growth of the consumption and, thus, errors that increase gradually; being the wavelets model with multiple regression a better alternative in this case.

From this point of view, the main contribution of this work is to provide a load forecasting model for decision support, applying the process of pattern extraction from the power consumption and estimating it, in order to establish more advantageous contracts of energy purchase in the future market for the power suppliers; especially given that the expansion of the power supply in the Amazonian region is a predominant factor of social development.

Acknowledgments. The authors would like to thank the power supplier CELPA for providing with the data for the analyzes. The work was also supported by the National Counsel of Technological and Scientific – CNPq.

References

1. Douglas, A.P., Breipohl, A.M., Lee, F.N., Adapa, R.: The impacts of temperature forecast uncertainty on Bayesian load forecasting. *IEEE Transactions on Power Systems*, 1507–1513 (1998)
2. Senjyu, T., Takara, H., Uezato, K., Funabashi, T.: One-hour-ahead Load Forecasting Using Neural Network. *IEEE Transactions on Power Systems*, 113–118 (2002)
3. Rocha, C., Santana, Á.L., Francês, C.R., Rego, L., Costa, J.: Decision Support in Power Systems Based on Load Forecasting Models and Influence Analysis of Climatic and Socio-Economic Factors. In: *Proceedings of SPIE Optics East*, vol. 6383 (2006)
4. ANEEL: Brazilian power systems atlas. In: *Agência Nacional de Energia Elétrica, Brasília – DF* (2003)
5. Pindyck, R.S., Rubinfeld, D.L.: *Econometric Models and Economic Forecasts*. Irwin/McGraw-Hill (1998)
6. Rice, J.A.: *Mathematical Statistics and Data Analysis*. Duxbury Press (1995)
7. Dillon, W.R., Goldstein, M.: *Multivariate Analysis - Methods and Applications*. John Wiley and Sons, Chichester (1984)
8. Adya, M., Collopy, F.: How Effective are Neural Networks at Forecasting and Prediction? A Review and Evaluation. *Journal of Forecasting*, 481—495 (1998)
9. Dohv, E., Feigin, P., Greig, D., Hyams, L.: Experience with FNN models for medium term power demand predictions. *IEEE Transactions on Power Systems*, 538–546 (1999)
10. Haykin, S.: *Neural Networks: a comprehensive Foundation*. Prentice-Hall, Englewood Cliffs (1998)
11. Moré, J.J.: The levenberg-marquardt algorithm: Implementation and theory. In: *Proceedings of Springer-Verlagin Numerical Analysis. Lecture Notes in Mathematics*, pp. 105–116 (1977)
12. Yu, P., Goldenberg, A., Bi, Z.: Time Series Forecasting using Wavelets with Predictor-Corrector Boundary Treatment. In: *Proceedings of the Temporal Data Mining Workshop at the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2001)
13. Daubechies, I.: *Ten Lecture on Wavelets*. CBMS series, Philadelphia (1992)
14. Zhang, Q., Benveniste, A.: Wavelet network. *IEEE Transactions on Neural Networks*, 889–898 (1992)

An Efficient and Robust Algorithm for Improving the Resolution of Video Sequences

Yubing Han, Rushan Chen, and Feng Shu

School of Electronic Engineering & Optoelectronic Techniques, Nanjing University of Science & Technology, Nanjing, Jiangsu 210094, China

Abstract. An efficient and robust super-resolution reconstruction algorithm for video sequences is proposed. In this algorithm, the L_1 and L_2 norms are introduced to form the data fusion term according to whether there exists motion estimation, and a robust Bilateral-TV regularization term is added to overcome the ill-posed problem of super-resolution estimation. Furthermore, we propose the use of regularization functional instead of a constant regularization parameter. The regularization functional is defined in terms of the reconstructed image at each iteration step, therefore allowing for the simultaneous determination of its value and the reconstruction of the super-resolution image. The iteration scheme, convexity and control parameter are thoroughly studied. Experimental results demonstrate the power of the proposed method.

Keywords: Super-Resolution Reconstruction, Bilateral-TV, Regularization, Euler-Lagrange Equation.

1 Introduction

In recent years, image and video super-resolution reconstruction has attracted increasing attention, and the applications can be widely founded in a broad range of image and video processing task such as aerial photo, medical imaging, video surveillance etc [1]. There are a variety of methods for super-resolution reconstruction such as Bayesian maximum a-posteriori, least square, projection onto convex sets and non-local-means etc [2-6]. Unfortunately, these methods are usually very sensitive to their assumed models of data and noise, which limits their utility. In [7], Farsiu et al. propose a robust algorithm to restore the super-resolution image using L_1 norm minimization and robust regularization based on a bilateral prior to deal with different data and noise models.

In this paper, an efficient and robust algorithm for improving the resolution of video sequences is proposed. In this algorithm, the L_1 and L_2 norms are adopted in data fusion term according to whether there exists motion estimation, which is more efficient to utilize the information at reference frame and is robust to the motion estimation error at other frames. On the other hand, a robust Bilateral-TV regularization term with simultaneous adaptation of regularization parameter is added to overcome the ill-posed problem of super-resolution reconstruction.

The paper is organized as follows. In section 2, we put forward the observation model of degraded images. Then the data fusion and regularization for super-resolution reconstruction are proposed in section 3 and section 4 respectively. The adaptive choice of regularization functional is presented in section 5. The iteration scheme is investigated in section 6. Experimental results are presented in section 7, and conclusions are drawn in section 8.

2 Observation Model

Supposed there is an observed low resolution video sequence, which is generated from the original high-resolution video sequence via following formula

$$Y_t = D_t B_t X_t + V_t, t > 0 \tag{1}$$

where t is the discrete time index, Y_t and X_t are the low-resolution and the high-resolution images represented in the lexicographic-ordered vector form, with a size of $L_1 L_2 \times 1$ and $M_1 M_2 \times 1$ respectively. D_t is decimation operation, B_t is space-invariant blur matrix, V_t stands for the random noise which are independent.

On the other hand, there is an equation representing the relationship of frames of video sequence, given by

$$X_k = F_{k,t} X_t + S_{k,t} \tag{2}$$

where $F_{k,t}$ is the motion compensation matrix which accounts for object motion occurring between frame X_t and X_k , $S_{k,t}$ is motion estimation error. Supposed t is reference time, we have

$$Y_k = A_{k,t} X_t + E_{k,t} \tag{3}$$

where $A_{k,t} = D_k B_k F_{k,t}$ represents the overall effect of decimation, blur convolution and motion compensation, $E_{k,t}$ represents the measured noise and motion estimation error.

3 Efficient and Robust Data Fusion

With the establishment in section 2, the goal of the super-resolution reconstruction is to produce the high-resolution image X_t based on a few low-resolution observations Y_k . Without loss of generality, we assume $t - N \leq k \leq t + N$, that is to say, the $2N + 1$ images centre in reference frame are adopted to restore the high-resolution image X_t . A popular family of estimation are the maximum likelihood type data fusion such that

$$\hat{\mathbf{X}}_t = \arg \min_{\mathbf{X}_t} \left\{ \sum_{k=t-N}^{t+N} \rho_k(\mathbf{Y}_k, \mathbf{A}_{k,t} \mathbf{X}_t) \right\} \tag{4}$$

where ρ_k is measuring the ‘‘distance’’ between the model and measurements. When the model error $\mathbf{E}_{k,t}$ is assumed as white Gaussian noise, the ρ_k is the L_2 norm of residual and the least squares formulation is achieved [3]

$$\hat{\mathbf{X}}_t = \arg \min_{\mathbf{X}_t} \left\{ \sum_{k=t-N}^{t+N} \|\mathbf{Y}_k - \mathbf{A}_{k,t} \mathbf{X}_t\|_2^2 \right\} \tag{5}$$

It is well known that the least squares estimation is a non-robust estimation when the data set contaminated with non-Gaussian outliers, and produces an image with visually apparent errors. To overcome the shortcomings of least squares estimation, the authors of reference [7] assume that the model error is the Laplacian distribution, the ρ_k is the L_1 norm of residual and the minimization problem becomes

$$\hat{\mathbf{X}}_t = \arg \min_{\mathbf{X}_t} \left\{ \sum_{k=t-N}^{t+N} \|\mathbf{Y}_k - \mathbf{A}_{k,t} \mathbf{X}_t\|_1 \right\} \tag{6}$$

From [7], we can show that the minimization problem based on L_1 norm is the most robust functional, and results in an approximate performance of pixel-wise median which cannot utilize all measured information fully. On the other hand, there is no need of motion estimation at reference frame and it is reasonable to assume that the model error $\mathbf{E}_{t,t}$ is Gaussian distribution. Therefore we can modify the minimization problem (6) by introducing the L_2 norm at reference frame such that

$$\hat{\mathbf{X}}_t = \arg \min_{\mathbf{X}_t} \left\{ \|\mathbf{Y}_t - \mathbf{A}_{t,t} \mathbf{X}_t\|_2^2 + \sum_{k \neq t} \|\mathbf{Y}_k - \mathbf{A}_{k,t} \mathbf{X}_t\|_1 \right\} \tag{7}$$

Obviously, this formation emphasizes the contribution of current reference frame and shows robust ability to the motion estimation error at other non-reference frames. So the performance based on such minimization is more robust than least square estimation based on L_2 norm, and is more efficient than the estimation based on L_1 norm.

4 Robust Regularization

Because super-resolution is an ill-posed problem, there exist an infinite number of solutions for the under-determined case or the solution for square and over-determined cases is not stable. Therefore, considering regularization in super-resolution algorithm as a means for picking a stable solution is very useful, if not necessary. Also, regularization can help the algorithm to remove artifacts from the final answer and improve the rate of convergence. Owing to the good property of preserving edges, here we introduce a robust regularizer called Bilateral-TV and modify the minimization problem as follow [7]

$$\hat{\mathbf{X}}_t = \arg \min_{\mathbf{X}_t} \left\{ \begin{aligned} & \left\| \mathbf{Y}_t - \mathbf{A}_{t,t} \mathbf{X}_t \right\|_2^2 + \sum_{k \neq t} \left\| \mathbf{Y}_k - \mathbf{A}_{k,t} \mathbf{X}_t \right\|_1 \\ & + \alpha \sum_{\substack{l=-P \\ l+m \geq 0}}^P \sum_{m=0}^P \gamma^{m+l} \left\| \mathbf{X}_t - \mathbf{S}_x^l \mathbf{S}_y^m \mathbf{X}_t \right\|_1 \end{aligned} \right\} \quad (8)$$

where $R(\mathbf{X}_t) = \sum_{\substack{l=-P \\ l+m \geq 0}}^P \sum_{m=0}^P \gamma^{m+l} \left\| \mathbf{X}_t - \mathbf{S}_x^l \mathbf{S}_y^m \mathbf{X}_t \right\|_1$ is regularization term and α is regularization parameter which is a scalar for properly weighting the data fitting term against the regularization term.

The matrices (operators) \mathbf{S}_x^l and \mathbf{S}_y^m shift \mathbf{X}_t by l and k pixels in horizontal and vertical directions respectively, presenting several scales of derivatives. The scalar weight γ is applied to give a spatially decaying effect to the summation of the regularization terms.

5 The Choice of Regularization Functional

In this section, we discuss the choice of regularization parameter α . When it becomes larger, the reconstructed image is blurrier, and when it becomes smaller, the reconstructed image is shown much noisier. How to decide regularization parameter is a very difficult and open problem. Currently, there have been a number of approaches developed to handle this problem, such as constrained least square (CLS), generalize cross validation (GCV) and L-Curve methods [8-10]. Unfortunately, all of these methods determine the regularization parameter in a separate first step, which either need additional computational overhead or need the prior knowledge about signal and noise. To overcome this difficulty, we modify regularization parameter to be a regularization functional [11, 12], such that

$$\alpha(\mathbf{X}_t) = \frac{D(\mathbf{X}_t)}{\frac{1}{\tau} - R(\mathbf{X}_t)} \quad (9)$$

where $D(\mathbf{X}_t) = \left\| \mathbf{Y}_t - \mathbf{A}_{t,t} \mathbf{X}_t \right\|_2^2 + \sum_{k \neq t} \left\| \mathbf{Y}_k - \mathbf{A}_{k,t} \mathbf{X}_t \right\|_1$ and τ is a control parameter.

Thus the minimization problem of super-resolution reconstruction becomes

$$\begin{aligned} \hat{\mathbf{X}}_t &= \arg \min_{\mathbf{X}_t} J(\mathbf{X}_t) \\ J(\mathbf{X}_t) &= D(\mathbf{X}_t) + \alpha(\mathbf{X}_t)R(\mathbf{X}_t) \end{aligned} \quad (10)$$

Obviously, the regularization functional has some properties as follows:

- (1) $\alpha(\mathbf{X}_t)$ is a smoothing functional with linear correlation of $\alpha(\mathbf{X}_t) = \tau J(\mathbf{X}_t)$.

(2) $\alpha(X_t)$ is a monotonically increasing functional with respect to $D(X_t)$ and $R(X_t)$. That is to say, if $D(X_t)$ is relatively large (meaning that the data fitting error is large), then a larger value of $\alpha(X_t)$ should be used. On the other hand, if $R(X_t)$ is relatively small (meaning that the image is more regular), then the smaller $\alpha(X_t)$ should be adopted, and the more details representing image singularity are further restored.

Now we discuss the choice the control parameter τ , whose requirement is to preserve the convexity of $J(X_t)$. Since a local extremum of a nonlinear convex functional becomes a global extremum, the iterative algorithm that will be employed for obtaining a minimizer of $J(X_t)$ will not depend on the initial condition. It was shown in [11] that

$J(X_t)$ is convex if the condition $\tau < \frac{1}{R(X_t)}$ is satisfied. For general image signal,

without loss of generality, it can be assumed that the signal energy $\|X_t\|^2$ is much larger than the singularity energy $R(X_t)$, that is to say, we can approximate $R(X_t) \leq \|X_t\|^2$. Since $\|X_t\|^2 \approx \|Y_k\|^2$, the control parameter can be selected as

$$\tau = \frac{1}{\max_k (\|Y_k\|)} \approx \frac{1}{\|X_t\|^2} < \frac{1}{R(X_t)} \tag{11}$$

and the condition for convexity is satisfied.

6 The Iterative Reconstruction Algorithm

As has already been mentioned, a super-resolution image is a minimum of the smoothing functional defined in (10). The corresponding Euler-Lagrange equation is

$$\frac{dJ(X_t)}{dX_t} = 0 \tag{12}$$

Differentiating $J(X_t)$ with respect to X_t , we obtain

$$\begin{aligned} 0 = & A_{t,t}^T (A_{t,t} X_t - Y_t) + \sum_{k \neq t} A_{k,t}^T \text{sign}(A_{k,t} X_t - Y_k) \\ & + \alpha(X_t) \sum_{\substack{l=-P \\ l+m \geq 0}}^P \sum_{m=0}^P \gamma^{m+l} (I - S_y^{-m} S_x^{-l}) \text{sign}(X_t - S_x^l S_y^m X_t) + R(X_t) \frac{d\alpha(X_t)}{dX_t} \end{aligned} \tag{13}$$

where S_x^{-l} and S_y^{-m} define the transposes of matrices S_x^l and S_y^m respectively and have a shifting effect in the opposite directions as S_x^l and S_y^m .

Since $\frac{d\alpha(\mathbf{X}_t)}{d\mathbf{X}_t} = \frac{d\alpha(\mathbf{X}_t)}{dJ(\mathbf{X}_t)} \frac{dJ(\mathbf{X}_t)}{d\mathbf{X}_t}$, when $\frac{d\alpha(\mathbf{X}_t)}{dJ(\mathbf{X}_t)} = \tau$ is bounded, it is clear that $\frac{d\alpha(\mathbf{X}_t)}{d\mathbf{X}_t} = 0$. Thus we obtain

$$\begin{aligned} & \mathbf{A}_{t,t}^T(\mathbf{A}_{t,t}\mathbf{X}_t - \mathbf{Y}_t) + \sum_{k \neq t} \mathbf{A}_{k,t}^T \text{sign}(\mathbf{A}_{k,t}\mathbf{X}_t - \mathbf{Y}_k) \\ & + \alpha(\mathbf{X}_t) \sum_{\substack{l=-P \\ l+m \geq 0}}^P \sum_{m=0}^P \gamma^{m+l} (\mathbf{I} - \mathbf{S}_y^{-m} \mathbf{S}_x^{-l}) \text{sign}(\mathbf{X}_t - \mathbf{S}_x^l \mathbf{S}_y^m \mathbf{X}_t) = 0 \end{aligned} \tag{14}$$

This is a very complex nonlinear equation, and it is solved by employing the successive iterative scheme, such that

$$\begin{aligned} & \mathbf{X}_t^{(n+1)} = \mathbf{X}_t^{(n)} \\ & - \beta \left[\begin{aligned} & \mathbf{A}_{t,t}^T(\mathbf{A}_{t,t}\mathbf{X}_t^{(n)} - \mathbf{Y}_t) + \sum_{k \neq t} \mathbf{A}_{k,t}^T \text{sign}(\mathbf{A}_{k,t}\mathbf{X}_t^{(n)} - \mathbf{Y}_k) + \alpha(\mathbf{X}_t^{(n)}) \\ & \sum_{\substack{l=-P \\ l+m \geq 0}}^P \sum_{m=0}^P \gamma^{m+l} (\mathbf{I} - \mathbf{S}_y^{-m} \mathbf{S}_x^{-l}) \text{sign}(\mathbf{X}_t^{(n)} - \mathbf{S}_x^l \mathbf{S}_y^m \mathbf{X}_t^{(n)}) \end{aligned} \right] \end{aligned} \tag{15}$$

where β is a scalar defining the step size in the direction of the gradient and n is iteration index.

7 Experimental Results

We use the test sequence ‘‘Caltrain’’ with size of 512×400, which can be founded at <http://cc.usu.edu/~arohb/caltrain.zip>. The original high resolution images are blurred by [5×5] Gaussian kernel with standard deviation 0.5, decimated using 2:1 decimation ration on each axis, and added by zero mean Gaussian white noise with standard deviation 2. At each time, $2N + 1 = 15$ images centre in current frame are adopted to reconstruct the super-resolution image. The simple block matching with exhaustive search has been used for motion estimation, the size of macro blocks is 16×16, and the search area is constrained up to 7 pixels on all fours sides of the corresponding macro block in reference frame. The regularization functional $\alpha(\mathbf{X}_t)$ is determined by (9) and the parameters $\gamma = 0.6$, $P = 2$ and $\beta = 0.1$. The iterative stop condition at

each frame is $\frac{\|\mathbf{X}_t^{(n)} - \mathbf{X}_t^{(n-1)}\|^2}{\|\mathbf{X}_t^{(n-1)}\|^2} \leq 10^{-8}$. In our experiment, the peak signal-to-noise

ratio (PSNR) is adopted for evaluating the reconstruction quality. For comparison, we also test the super-resolution reconstruction methods based on L_2 and L_1 norms in

the data fitting term respectively, the regularizer is also Bilateral-TV. For convenience, we call them “L2+BTv” and “L1+BTv” methods, and the super-resolution algorithm proposed in this paper is named as “L2+L1+BTv” method.

Fig.1 and Fig.2 give the super-resolution reconstruction results and the curves of PSNR-Frame of 15th frame in “Caltrain” sequence using different super-resolution methods. From these results, we can see that:

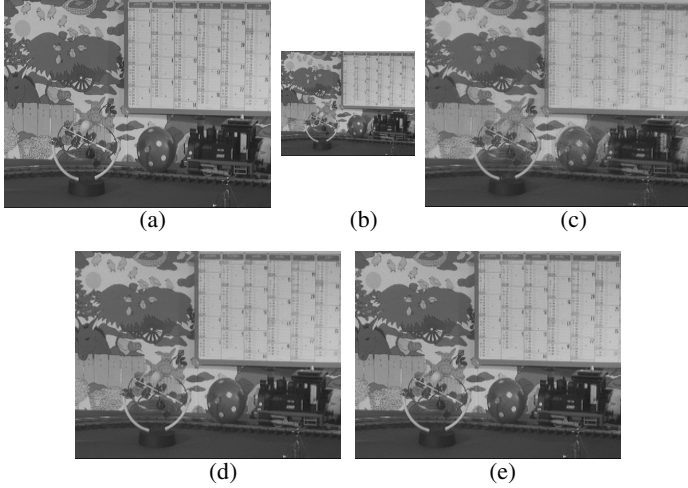


Fig. 1. The super-resolution results of 15th frame in Caltrain sequence (a) original image, (b) degraded image, (c) L2+BTv, (d) L1+BTv, (e) L2+L1+BTv

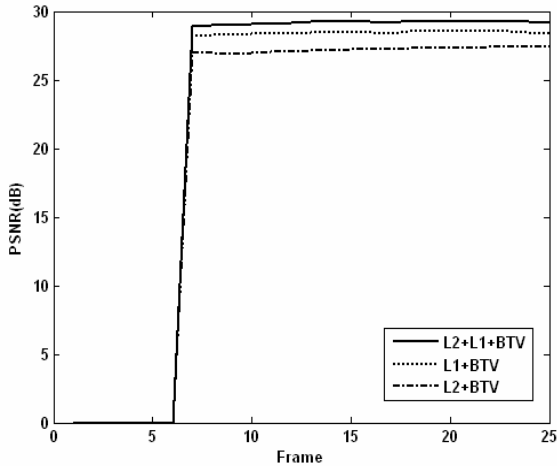


Fig. 2. The curves of PSNR-Frame using different super-resolution reconstruction methods

- (1) Compared with “L2+BTv” and “L1+BTv”, the “L2+L1+BTv” method can obtain some improvement in PSNR. The performance of method “L2+BTv” is worst, and the performance of “L1+BTv” is better than “L2+BTv”.
- (2) In visual inspection, the reconstruction image of “L2+BTv” method is shown noisier in motion areas because of serious motion estimation error; whereas the result of “L1+BTv” method is shown noisier in motionless areas because of no full fusion of all measured information. The “L2+L1+BTv” method can take advantage of “L2+BTv” and “L1+BTv”, and has a middle reconstruction effect between them.

8 Conclusions

In this paper, an efficient and robust algorithm is proposed for video super-resolution reconstruction. After discussed the shortcomings of estimation based on L_2 and L_1 norms, a new data fusion term with L_2 norm at reference frame and L_1 norm at other non-reference frames is introduced, which is not only very robust to the modeling error of motion estimation, but also is very efficient to utilize the information of reference frame. Secondly, a robust regularization term called Bilateral-TV is added to overcome the ill-posed problem of super-resolution estimation, which results in reconstructed image with sharp edges. Then an adaptive choice of regularization functional is presented, which is determined by the restored image at each step. Finally, the numerical iteration, convexity and control parameter are thoroughly investigated. Experimental results indicate that the proposed algorithm is very effective when exiting the inter-frame motion estimation error.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (No. 60802039) and Jiangsu Planned Projects for Postdoctoral Research Funds (No. 0702023B).

References

1. Park, S.C., Pak, M.K., Kang, M.G.: Super-resolution Image Reconstruction: a Technical Overview. *IEEE Signal Processing Magazine* 20(3), 21–36 (2003)
2. Nguyen, N., Milanfar, P., Golub, G.: A Computationally Efficient Superresolution Image Reconstruction Algorithm. *IEEE Transactions on Image Processing* 10(4), 573–583 (2001)
3. Elad, M., Hel-Or, Y.: A Fast Super-resolution Reconstruction Algorithm for Pure Translational Motion and Common Space Invariant Blur. *IEEE Transactions on Image Processing* 10(8), 1187–1193 (2001)
4. Shen, H., Zhang, L., Huang, B., et al.: A MAP Approach for Joint Motion Estimation, Segmentation, and Super Resolution. *IEEE Transactions on Image Processing* 16, 479–490 (2007)
5. Altunbasak, Y., Patti, A.J., Mersereau, R.M.: Super-resolution Still and Video Reconstruction from MPEG-coded Video. *IEEE Transactions on Circuits and Systems for Video Technology* 12, 217–226 (2002)

6. Protter, M., Elad, M., Takeda, H., Milanfar, P.: Generalizing the Non-Local-Means to Super-Resolution Reconstruction. *IEEE Transactions on Image Processing* (to appear, 2008)
7. Farsiu, S., Robinson, D., Elad, M., Milanfar, P.: Fast and Robust Multi-frame Super-Resolution. *IEEE Transactions on Image Processing* 13(10), 1327–1344 (2004)
8. Banham, M.R., Katsaggelos, A.K.: Digital Image Restoration. *IEEE Signal Processing Magazine* 14(2), 24–41 (1997)
9. Golub, G.H., Heath, M., Wahba, G.: Generalized Cross-validation as a Method for Choosing a Good Ridge Parameter. *Technometrics* 21(2), 215–223 (1979)
10. Bose, N.K., Lertrattanapanich, S., Koo, J.: Advances in Superresolution using L-curve. In: *The IEEE International Symposium on Circuits and Systems*, Sydney, Australia, vol. 2, pp. 433–436 (2001)
11. Kang, M.G., Katsaggelos, A.K.: General Choice of the Regularization Functional in Regularized Image Restoration. *IEEE Transactions on Image Processing* 4(5), 594–602 (1995)
12. He, H., Kondi, L.P.: Resolution Enhancement of Video Sequences with Simultaneous Estimation of the Regularization Parameter. *SPIE Journal of Electronic Imaging* 13, 586–596 (2004)

Research on Variable Step-Size Blind Equalization Algorithm Based on Normalized RBF Neural Network in Underwater Acoustic Communication

Xiaoling Ning, Zhong Liu, and Yasong Luo

Electronics Engineering College,
Naval University of Engineering , Wuhan 430033, China

Abstract. In this paper, based on constant modulus algorithm (CMA), variable step-size blind equalization algorithm based on normalized radial basis function (RBF) neural network is proposed, considering blind equalization can equalize nonlinear characteristic of underwater acoustic channel without training sequence and RBF neural network is a nonlinear system with excellent approximation characteristic and performance of equalizing nonlinear channel. The algorithm is emulated in SIMULINK and verified its feasibility and performance using data of lake testing. Simulation and testing results show that variable step-size blind equalization algorithm based on normalized RBF neural network is better than classical BP algorithm and RBF algorithm in convergence rate and equalization performance.

Keywords: Underwater acoustic channel; RBF neural network; blind equalization; higher-order squared error.

1 Introduction

With a good many factors such as multipath effects and reverberation, it is necessary that underwater acoustic channel has nonlinear characteristics. Neural network is a dynamic nonlinear system, which has prodigious application potential, the studies applying neural network to perform underwater acoustic equalization are increasing [1]. BP neural network—clear, simple and active state steady—is applied widely, but its learning speed is very slow and utility is limited in high real-time situation for it is a global approximation neural network. RBF network is a local approximation neural network, which only needs to correct a small quantity of weights and threshold values; and its learning speed is quite fast. In 1999, Chen and other researchers began to study RBF neural network equalizers and some algorithms have been proposed [2]-[4]. Though RBF network equalizer is better than BP network equalizer in convergence performance, its equalization performance is not exerted fully. To make the effect of equalization better like faster convergence rate and smaller mean-squared error (MSE), this paper presents a variable step-size blind equalization algorithm based on normalized RBF neural network.

This paper is organized as follows. Section 2 briefly presents the model of blind equalization based on neural network. In Section 3, the structure of RBF network is presented and the theoretical analyzes of step-size and weights of the proposed

algorithm are induced. Section 4 reports the simulation results. Section 5 shows the application of the algorithm. Conclusions are given in Section 6.

2 Model of Blind Equalization Algorithm Based on Neural Network

The basic idea of blind equalization based on neural network is that neural network replaces transversal filter in classical constant module algorithm, using cost function of selected blind equalization algorithm to modulate connection weights, making output sequence approach to sending sequence. The schematic diagram is shown in fig. 1 [5].

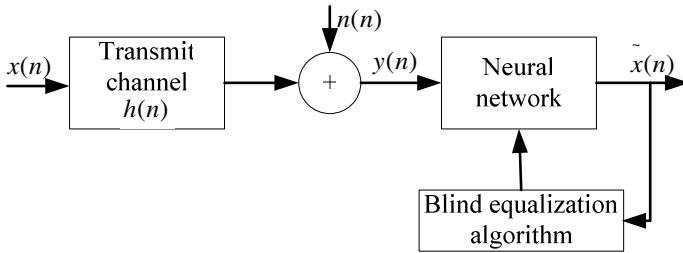


Fig. 1. The schematic diagram of blind equalization based on neural network

3 Structure Model and Algorithm of Radial Basis Function Equalizer

3.1 Normalized RBF Neural Network Model

RBF neural network is a forward feedback neural network, which has two network layers, hidden layer is RBF layer, and output layer is linear layer. As seen from the function approximation, the principle of RBF network is that any function can be expressed weighing sum of a group of primary functions if network is seen as approximation of a unknown function [6]. The structure of RBF network equalizer is shown in fig. 2.

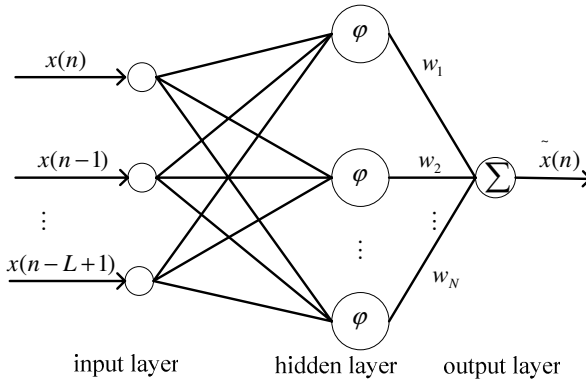


Fig. 2. Elementary principle of target depth estimation

The input of network is signal vector $X(n) = [x(n), x(n-1), \dots, x(n-L+1)]^T$, namely, it is L delays of aberrant complex signal $X(n)$. RBF function $\varphi(n)$ is gauss function. Network output $\tilde{x}(n)$ is that

$$\tilde{x}(n) = \sum_{i=1}^N w_i \varphi_i(n) \tag{1}$$

In expression (1), normalized RBF $\varphi_i(n)$ is defined as

$$\varphi_i(n) = G(\|x - t_i\|^2) = \exp\left[-\frac{m_1}{d_{\max}^2} \|x - t_i\|^2\right] = \exp\left[-\frac{\|x - t_i\|^2}{2\sigma_i^2}\right], \quad i = 1, 2, \dots, m_1 \tag{2}$$

$t_i (1 \leq i \leq L)$ is complex center vector, m_1 is center numbers, d_{\max} is the maximal distance among centers. We note that all standard deviations of gauss RBF are fixed as

$$\sigma = \frac{d_{\max}}{\sqrt{2m_1}} \tag{3}$$

Above formula ensure every RBF is neither too tip nor too even (lubricity is good).

3.2 Higher-Order Squared Error Modulating Learning Step-Size

According to constant module algorithm [7], cost function is that

$$J_D = \frac{1}{2} \left[\left| \tilde{x}(k) \right|^2 - R_{CM} \right]^2 \tag{4}$$

Where

$$R_{CM} = \frac{E\left(\left|\tilde{x}(k)\right|^4\right)}{E\left(\left|\tilde{x}(k)\right|^2\right)} \tag{5}$$

Higher-order squared error can reflect the maximal and minimum phase component, which is accordant with the nonlinear characteristic of neural network. Integrating CMA and output of RBF network, error function is that

$$e(k) = \tilde{x}(k) \left(\left| \tilde{x}(k) \right|^2 - R_{CM} \right) \tag{6}$$

From (6) we observe that error is related to inputs and weight coefficients of equalizer.

Because of various interference, the instantaneous value of $e(k)$ is unstable, it can be seen as a random variable. The digital statistic characteristics of random variable

can intensively reflect some average characteristics, so this paper uses higher-order squared error as step-size modulating gene to realize step-size modulation.

The m -order center moment of error is that

$$C_m(k) = E[e(k) - \bar{E}]^m \quad (7)$$

Where, \bar{E} is mean of error $e(k)$

$$\bar{E} = \frac{1}{L+1} \sum_{k=0}^L e(k) \quad (8)$$

In the application, we need do random processing for received signal, making it to meet independent identically distributed characteristic. Therefore, this paper supposes that the studied signal has symmetrical probability density function; its odd number order moment is zero. Thereby, (7) can be written as

$$C_m(k) = E[e(k)]^m - \bar{E}^m \quad (9)$$

Step-size modulating method is that

$$\mu(k+1) = \mu(k) + [C_m(k+1) - C_m(k)] \quad (10)$$

In order to ensure the solidity of algorithm, step-size is restricted according to adaptive transversal filter.

Namely

$$0 < \mu(k) < \frac{1}{tr[R]} \quad (11)$$

Note that R is autocorrelation matrix of equalizer inputs.

3.3 Learning Step-Size Modulates Network Weights

In term of learning step-size $\mu(k)$, the iterative expressions of network weights is that

$$w(k+1) = w(k) - \mu(k) \frac{\partial J_D}{\partial w(k)} \quad (12)$$

$$\frac{\partial J_D}{\partial w(k)} = 2\{x^2(k) - R_{CM}\} \tilde{x}(k) \frac{\partial \tilde{x}(k)}{w(k)} \quad (13)$$

Where

$$\frac{\partial \tilde{x}(k)}{\partial w_i(k)} = \varphi_i(n) \quad (14)$$

Combining (8) and (13), we note that variable step-size algorithm adopts bigger step-size at the beginning, quickening learning speed of weights w_i . When algorithm gradually is becoming convergent, especially algorithm astringes to be close to global CMA minima, learning step-size is reduced to achieve higher convergence precision.

The tenet of choosing m value is that if signal-to-noise ratio (SNR) is bigger, we choose a bigger m value; if SNR is smaller, we properly reduce m value to prevent deviation of higher-order squared error estimation pricking up as a result of noise disturbing, which possibly leading algorithm diverging.

4 Simulation

In simulation, we choose the simplest binary equiprobable sequence, and modulation system is BPSK in the absence of zero –mean tape-limiting white noise. In order to embody the equalization performance of Variable Step-size Blind Equalization Algorithm Based on Normalized RBF Neural Network for nonlinear channel, the adopted nonlinear channel model of the simulation is that

$$h(n) = x(n) + 0.2x^2(n) \tag{15}$$

Where

$$x(n) = s(n) + 0.5s(n-1) \tag{16}$$

The initial values of mean-square deviation are 1, center vector initial values of hidden node are N input vectors at the beginning, the initial values of weight coefficients are the random numbers. By simulating, we obtain fig. 3 and fig. 4. In fig. 3 and fig. 4, we use new RBF to substitute for variable step-size blind equalization algorithm based on normalized RBF neural network.

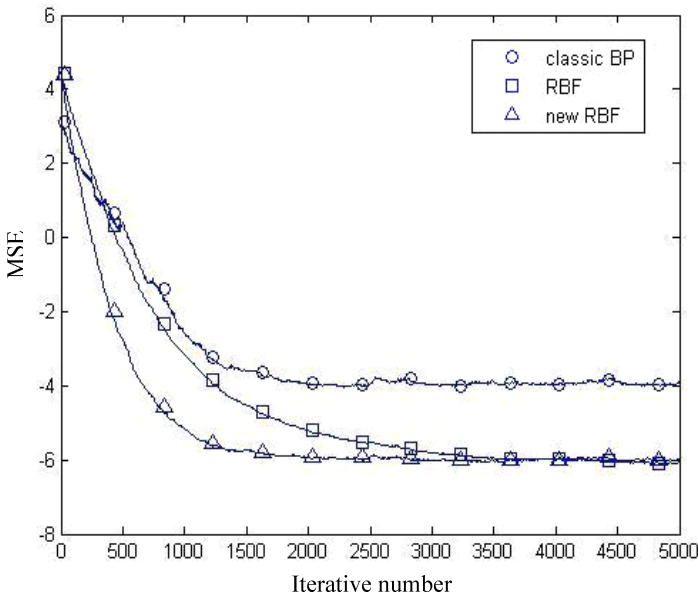


Fig. 3. The figure of learning curve for nonlinear channel, (SNR=20 db, $m = 4$)

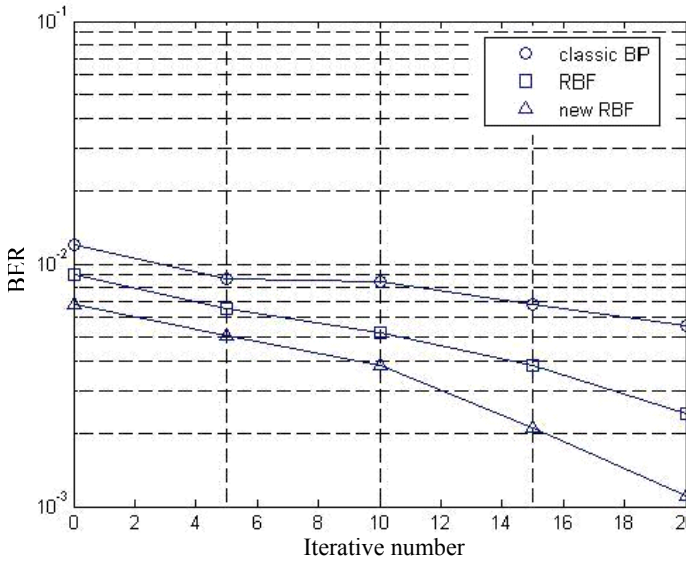


Fig. 4. Curve of BER

In fig. 3, we note that classical BP algorithm, RBF algorithm and variable step-size blind equalization algorithm based on normalized RBF neural network have different learning curves, at the same time, we also observe that variable step-size blind equalization algorithm based on normalized RBF neural network has faster rate of convergence and smaller steady-state residual error. Fig. 4 also show that variable step-size blind equalization algorithm based on normalized RBF neural network has lower error rate and better equalization performance.

5 Lake Testing and Digital Processing

Correspondingly, the sending signal of lake testing is BPSK, which meets the demand of blind equalization for sending data statistical characteristic. In testing, we examine the effect of variable step-size blind equalization algorithm based on normalized RBF neural network for coherent signal with multipath interference.

The transmission speed of data is 4Kbit/s and 2Kbit/s respectively in the testing, emitter and energy converter are placed at 5m under water, the distance between sender and receiver is 20m, 120m, 200m and 330m respectively.

Fig.5 and fig.6 are respectively the error curve of convergence and the figure of equalization output along with iterative course of data processing on condition that the distance between sending and receiving is 330m and the transmission rate is 2Kbit/s.

From fig.5 and fig.6, we note that convergence inclines to stabilization and the effect of convergence is very good when iterative number is up to 3000. Thus it can be seen that it is feasible that variable step-size blind equalization algorithm based on normalized RBF neural network is used in lake underwater communication.

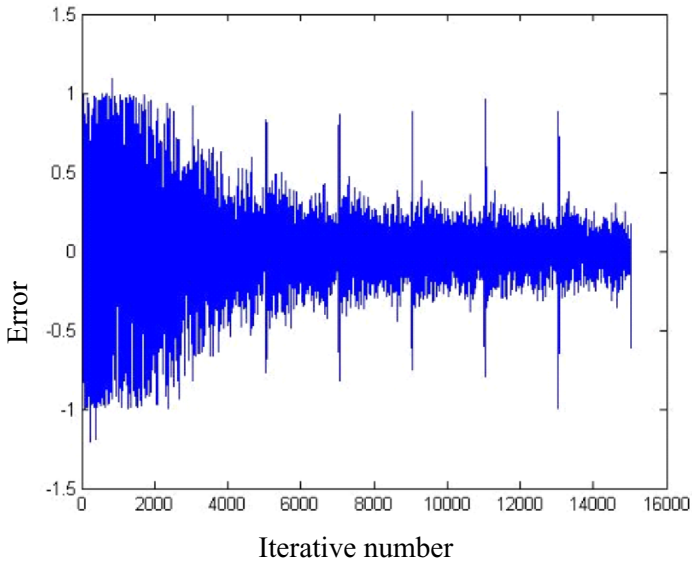


Fig. 5. Error curve of conver

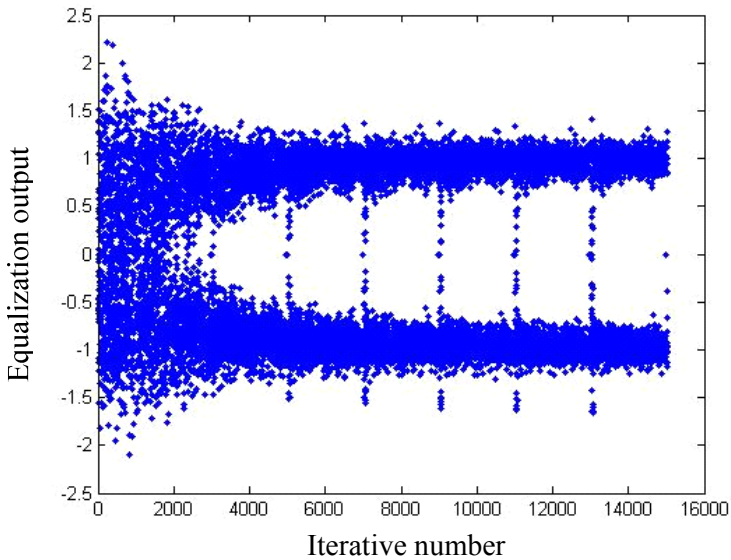


Fig. 6. Equalization output along with iterative course

6 Conclusion

RBF neural network which is used in underwater communication has unique predominance, namely, it has the local approximation characteristic. The result of

simulation and testing show that new algorithm has faster rate of convergence and better equalization performance by adjusting step-size. But higher-order squared error needs a great deal calculation; we should study how to decrease calculation capacity. In this paper, we only study BPSK, the non-CM M-QAM is still needed to study.

References

1. Guo, Y.C., He, L.Q., Han, Y.G., Zhao, J.W.: Blind Equalization Technology and Its Application in Underwater Acoustic Channel Equalization. *Ship Science and Technology* 29, 25–26 (2007)
2. Chen, S., Gibson, G.J., Cowan, C.F., et al.: Reconstruction of Binary Signals Using an Adaptive Radial Basis Function Equalizer. *Signal Processing* 22, 77–93 (1991)
3. Chen, S., Mclaughlin, S., Mulgrew, B.: Complex Valued Radial Basis Function Networks: Network Architecture and Learning Algorithm (Part I). *Signal Processing* 35, 19–31 (1994)
4. Chen, S., Mclaughlin, S., Mulgrew, B.: Complex Valued Radial Basis Function Networks: Application to Digital Communications Channel Equalization (Part II). *Signal Processing* 36, 175–188 (1994)
5. Yuan, L., Zhang, L.Y.: Blind Equalization Algorithm Based on Neural Network and Genetic Algorithm. *Journal of Taiyuan University of Technology*, 37–69 (2006)
6. Zhou, K.L., Kang, Y.H.: Neural network model and MATLAB simulation program design. Publishing house of Tsinghua University, Beijing (2005)
7. Johnson Jr., C.R., Schniter, P., Endres, J.T., et al.: Blind Equalization Using the Constant Modulus Criterion: A Review. *Proceeding of The IEEE* 86, 1944–1945 (1998)

The Analysis of Aircraft Maneuver Efficiency within Extend Flight Envelop

Hao Long¹ and Shujie Song²

¹ College of Automatics, Beijing Union University, Beijing 100101, China

² Aviation Corporation of China, Beijing 100712, China

Abstract. Using thrust vector must be one of key technique characters for the forth generation fighters especially in condition of High AOA(Angle of attack). The general effectors lost the maneuver efficiency or get poorly, so new type of control law must be found to improve the maneuver efficiency of aircraft within extended flight envelop. The control law should not only overcome those problems but also satisfy requirement of the forth generation aircraft. For those reasons, the Receding Horizon Optimal (RHO) control law is brought out to solve those problems. Simulation results comparison between using the general effectors and using the general effectors with thrust vector all together show: within the extend flight envelop thrust and thrust vector use together can guarantee aircraft following pilot command exactly, the RHO control law can solve those problems perfectly and satisfy the requirement of the forth generation fighter.

Keywords: Maneuver Efficiency, RHO, Flight Control, Flight Envelop.

1 Introduction

The perfect maneuver efficiency in extend flight envelop will be the symbol of forth generation fighter, which can ensure the fighter have excellence campaign capability. Fighter with thrust vector can enhance the performance of fighter greatly, not only within extend flight envelop, but also improve the maneuver efficiency of short takeoff or vertical land, such as F-16/MATV, X-31 were experimented to test stall flight and simulate wrestle between two fighters, which show that thrust vector is very useful [1,2]. But fighter with thrust vector may meet more complex problem about nonlinear aerodynamic coupling under stall maneuver and high AOA than those under general conditions. Therefore it is very difficult for aircraft to be controlled in the extreme condition by using the PID control systems. The PID control systems only overcome little disturbance. So a new type of control systems needed be searched to satisfy the control capability in case of stall maneuver and high AOA. These control system also should overcome the nonlinear coupling of aerodynamic.

Under these problems, this paper brings forward Receding Horizon Optimal algorithm and design the control law for aircraft within the extend flight envelop. The thrust vector influence to aircraft is analyzed when it flies under the condition of stall maneuver and high AOA.

2 Affine Nonlinear Movement Equations of Aircraft with Thrust Vector

Now, an affine nonlinear equation of aircraft with twin thrust vectors is described by:

$$\dot{x}(t) = f(x(t)) + e[x(t)] \cdot x(t) + g[x(t)] \cdot \delta(t) + h(x(t), \delta_T(t)) . \quad (1)$$

where state variables $x = [\alpha, \omega_z, \beta, \omega_x, \omega_y] \in R^{5 \times 1}$ respectively denote the AOA, angle rate of pitch, the angle of sideslip, angle rate of roll, and angle rate of row. General effectors input variables $\delta = [\delta_z, \delta_x, \delta_y]^T \in R^{3 \times 1}$ respectively denote elevators, ailerons, and rudders. $\delta_T = [\theta_{TL}, \theta_{TR}, \varphi_{TL}, \varphi_{TR}, P]^T \in R^{5 \times 1}$ denote thrust vector vanes and thrust value input. However, θ_{TL} and θ_{TR} respectively denote the angle between left or right thrust vector vane and xoz plane of fighter body coordinate system, φ_{TL} and φ_{TR} respectively denote the angle between left or right thrust vector vane and xoy plane of fighter body coordinate system, and P denotes thrust; $f(x(t))$ denotes the nonlinear influence of gravity and coupling items of fly state variables on the movement of fighters. $e[x(t)] \in R^{5 \times 5}$ denote the derivative matrices of aerodynamic, which can be achieved via interpolating with the database of aerodynamic saved in fighters; $g[x(t)] = \partial x / \partial u \in R^{5 \times 3}$ denote the derivative matrices of maneuver of general effectors, which is the nonlinear function of fly states; $h(x(t), \delta_T(t)) \in R^{5 \times 4}$ denote the projective relation of thrust vector on state variables of fighters, which is the nonlinear function of structure of aircraft and fly state variables and thrust vector.

3 Receding Horizon Optimal Control

RHO (Receding Horizon Optimal control) is derived from the process control and industry, it is an optimal algorithm of continue time zone in the model forecast control law. At the immediate time t , the horizon of linear quadratic is fixed up as T , and $t+T$ denote the end time. Suppose that the end state variables satisfy $x(t+T) = 0$. A series of control signal within the time section of $[t, t+T]$ can be resolved out through calculating the optimal linear quadratic algorithm of the model described as following, The control value at the immediate time t in the series control signal is chosen as the control input. At the next time t' , the control signal is not the next value in the series control signal at the time t , but the value at the time t in a new series control signal resolved from the optimal linear quadratic algorithm of the model following within $[t', t'+T]$. So the cycle calculating action result in a series of optimal control signal input, which is defined to be the optimal linear quadratic algorithm of models following^[3-6].

At the immediate time t , the nonlinear item $f(x(t))$ and $h(x(t), \delta_T(t))$ in (1) are dealt with 1st Taylor series spread for the state variables and thrust vector as: (subscript “0” denote the state at the immediate t).

$$f(x) \approx \left. \frac{\partial f(x)}{\partial x} \right|_{x=x_0} \cdot x + f(x_0) - \left. \frac{\partial f(x)}{\partial u} \right|_{x=x_0} \cdot x_0 \quad (2)$$

$$h(x, \delta_T) \approx \left. \frac{\partial h(x, \delta_T)}{\partial x} \right|_{\substack{x=x_0 \\ \delta_T=\delta_{T_0}}} \cdot x + \left. \frac{\partial h(x, \delta_T)}{\partial \delta_T} \right|_{\substack{x=x_0 \\ \delta_T=\delta_{T_0}}} \cdot (\delta_T - \delta_{T_0}) + h(x_0, \delta_{T_0}) - \left. \frac{\partial h(x, \delta_T)}{\partial x} \right|_{\substack{x=x_0 \\ \delta_T=\delta_{T_0}}} \cdot x_0 \quad (3)$$

according to (1),(2)and(3), the form of const linearization can be got at immediate time t below:

$$\dot{x}(t) = A_0 \cdot x(t) + B_{\delta_0} \cdot \delta(t) + B_{\delta_{T_0}} \cdot \Delta\delta(t) + d_0 \quad (4)$$

Therefore,

$$\Delta\delta_T(t) = \delta_T(t) - \delta_{T_0},$$

$$A_0 = \left. \frac{\partial f(x(t))}{\partial x(t)} \right|_{x(\tau)=x_0} + e[x(t)] \Big|_{x(\tau)=x_0} + \left. \frac{\partial h(x(t), \delta_T(t))}{\partial x(t)} \right|_{\substack{x(\tau)=x_0 \\ \delta_T(\tau)=\delta_{T_0}}} \in R^{5 \times 5}$$

$$B_{\delta_0} = g[x(t)] \Big|_{x(\tau)=x_0} \in R^{5 \times 3}, \quad B_{\delta_{T_0}} = \left. \frac{\partial h(x(t), \delta_T(t))}{\partial x(t)} \right|_{\substack{x(\tau)=x_0 \\ \delta_T(\tau)=\delta_{T_0}}} \in R^{5 \times 3}$$

$$d_0 = f(x_0) - \left. \frac{\partial f(x(t))}{\partial x(t)} \right|_{x(\tau)=x_0} x_0 + h(x_0, \delta_{T_0}) - \left. \frac{\partial h(x(t), \delta_T(t))}{\partial x(t)} \right|_{\substack{x(\tau)=x_0 \\ \delta_T(\tau)=\delta_{T_0}}} x_0 \in R^{5 \times 4}$$

The effectors through actuators influence the flight states of fighter. When aircraft tracks the reference signal tightly, it must have the ability to overcome the variation of models parameters and the resistance of disturbance robustly. However, the saturation of actuators would not ensure the aircraft track reference signal exactly. Then, the model of actuators must be considered in the maneuver of fighter. The general model of actuators is described as below:

$$\dot{\delta}(t) = A_\delta \delta(t) + B_\delta \delta_c(t), \quad (5)$$

where $\delta_c = [\delta_{zc} \quad \delta_{xc} \quad \delta_{yc}]^T$ denotes the command signal of actuators and $A_\delta \in R^{3 \times 3}$, $B_\delta \in R^{3 \times 3}$ denote the steady matrix and control matrix of actuators respectively.

RHO control algorithm need models to forecast the state of fighter and track it. Here using flight quality models provide the ideal models to be tracked. Because the complex flight quality models can be modeled by the simple 1st or quadratic models, which use the low rank systems to simulate the high rank systems best of all according to the optimal control theory. The flight quality models can be expressed as follows:

$$x_m = A_m x_m + B_m F_c, \tag{6}$$

where $x_m = [\omega_{zm} \quad \omega_{xm} \quad \omega_{ym}]^T$ denotes the command signal of three angle rates, which is resolved from the flight quality models. $A_m \in R^{3 \times 3}$ and $B_m \in R^{3 \times 3}$ denote the steady matrix and control matrix of the flight quality models respectively.

$F_c = [F_{zc} \quad F_{xc} \quad F_{yc}]$ denotes the three-axes pilot command.

The expanded linearization flight equations are derived from (4-6) as:

$$\dot{\bar{x}} = A\bar{x} + B\bar{u} + B_m F_c + d, \tag{7}$$

Where $\bar{x} = [x \quad \delta \quad x_m]^T \in R^{1 \times 1}$ denotes the states of the expanded linearization flight equation and $u = [\Delta\delta_r \quad \delta_c]^T \in R^{7 \times 1}$ denotes the input of them. The steady matrix A and the control matrix B of the expanded linearization flight equations are shown as below. (Note: the letter ‘‘O’’ denote zero matrixes)

$$A = \begin{bmatrix} A_0 & B_{\delta_0} & 0 \\ 0 & A_\delta & 0 \\ 0 & 0 & A_m \end{bmatrix} \in R^{11 \times 11}, B = \begin{bmatrix} B_\delta & 0 \\ 0 & B_{\delta_r} \end{bmatrix} \in R^{11 \times 7}$$

$d = [d_0 \quad 0]^T \in R^{1 \times 1}$ represents the remnant items. Four indexes of optimal performance are limited respectively for the following state errors $e = x_\omega - x_m$, the following state errors rate $\dot{e} = \dot{x}_\omega - \dot{x}_m$, the displacement of actuators states δ and the angle rate of actuators states, so the following servo optimal index of the RHO control is defined as below: (here the weighted $Q_e, Q_{\dot{e}}, Q_\delta, Q_{\dot{\delta}}, R$ are positive dialog matrices)

$$J = \|e\|_{Q_e}^2 + \|\dot{e}\|_{Q_{\dot{e}}}^2 + \|\delta\|_{Q_\delta}^2 + \|\dot{\delta}\|_{Q_{\dot{\delta}}}^2 + \|\bar{u}\|_R^2. \tag{8}$$

Because the states \bar{x} of the expend linear equations integrate the flight states and the states x_m of the flight quality models, the following servo problem defined by (8)

is changed to the linear regulating problem of system, thus the optimal algorithm is predigested. Adopting the Hamiltonian algorithm, the solutions and Riccati equations of the optimal control indexed by (8) is shown as:

$$\begin{cases} \bar{u} = M_1^{-1} \cdot (-P \cdot \bar{x} - M_2 \cdot s + M_3 \cdot F_c + M_4 \cdot d) \\ \dot{P} + PM_5 + M_5^T P - PBM_1^{-1}B^T P + M_6 = O \\ \dot{s} + M_7 \cdot s + M_8 \cdot F_c + M_9 \cdot d = O \end{cases} \quad (9)$$

The integrate section is limited within $[0, T]$; \bar{u} denote the solutions at the immediate time t ; P and s denote the Riccati matrix; the end conditions of RHO control is $P(T) = O^{1 \times 11}, s(T) = O^{1 \times 1}$; due to the expression of $M_i, i = 1 \dots 9$ within above three equations are very complex, only the modalities of them are shown above because of the length limit of this paper.

4 Simulation

The simulation validates thrust vector and thrust value affect the maneuver capability in condition of stall and high AOA. When aircraft flies within the extend flight envelop, the general effectors can't provide enough force of aerodynamic because control efficiency become low, In this condition, using thrust vector together with thrust value, the aircraft will produce the enough force and moment of aerodynamic, thus ensure the aircraft follow the reference state.

The simulation step $\Delta t = 0.0125s$. The initial flight condition is defined as that the height is 8 km and the Mach number is 0.1 within the extend flight envelop. The initial flight condition is set at $\alpha_0 = 30^\circ, \delta_{z0} = -18^\circ$, and other initial states and control variables are set as zero. For comparison the influence on the maneuver capability of fighter with or without thrust vector clearly. In the simulation the regulating process of engine is omitted. Engines only run in two states: the max thrust state and the full afterburner state. At the height of 8km, the max thrust of twin engines can achieve about 30000N, and the full afterburner of twin engines can achieve about 40000N.

The following section gives the different control results according to with or without thrust vector using RHO control theory.

Example 1. Simulation result when only using the general effectors, and engines working at the max thrust state

Twin engines work at the max thrust state 30000N. The positive gain matrices $Q(t)$ need to be regulated according to the importance of different optimal index, the control capacity of effectors, and the change of states. Certain gain matrices $Q(t)$ are chosen as:

$$Q_e(t) = \text{diag}(20, 60, 180), Q_e(t) = \text{diag}(0.01, 0.3, 10)$$

$$Q_\delta(t) = \text{diag}(1000, 1000, 1000), Q_\delta(t) = \text{diag}(5, 10, 5)$$

Because no thrust vector is used and no thrust value changed, only the general effectors is used within this condition, the objective gain matrix $R(t)$ is set as:

$$R(t) = \text{diag}(100, 10^4, 10^4, 10^4, 2.5, 2, 0.5)$$

The gain matrices corresponding to $P, \theta_{TL}, \theta_{TR}, \varphi_T$ lie in the former four lines of the dialog matrix $R(t)$. For keeping the const thrust value, the gain value corresponding to P is always set as 100. Duo to no effect of thrust vector being considered, the gain values corresponding to $\theta_{TL}, \theta_{TR}, \varphi_T$ are set up to the quantity 10000 to omit their effects. Three gain values corresponding to the general effectors are set as 2.5, 2 and 0.5 to increase the active capacity of the general effectors.

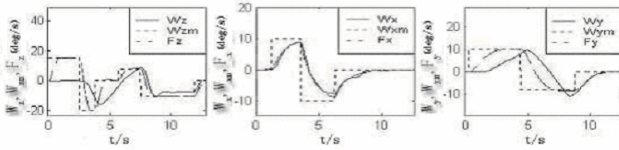


Fig. 1. The simulation result of pilot command, flight state, and flight quality model

In figure 1, three solid lines show the tracks of three angle rates which are the state ω_z of longitudinal (left figure), the state ω_x of lateral (middle figure) and the state ω_y of course (right figure); F_z (left figure), F_x (middle figure) and F_y (right figure) shown in dot step lines denote three pilot command for pitch, roll and row directions; three states following command ω_{zm} (left figure), ω_{xm} (middle figure) and ω_{ym} (right figure) shown in dash lines denote the calculating results of the flight quality which make pilot command as inputs.

From figure 1, it can be seen that only the state ω_x can follow the state command ω_{xm} tightly, and the flight states ω_z and ω_y can't follow the state command ω_{zm} and ω_{ym} . The reason is that the general effectors can't provide enough force and moment of aerodynamic to change the flight states to follow state commands in the condition of $8km$ height and $0.1Ma$.

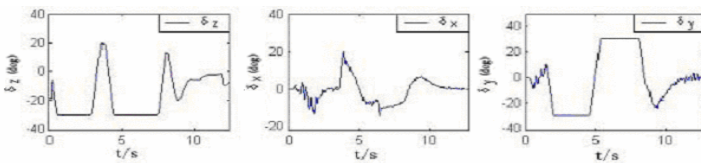


Fig. 2. The angle of elevator, aileron and rudder

In figure 2, the left figure, middle figure and right figure respectively denote the deflection of elevator, aileron and rudder. However, the deflection of elevator and rudder quite a lot of time are up to the saturation state 30° to satisfy following states commands and the deflection of aileron also move within a larger angle section $[-18^\circ, 21^\circ]$.

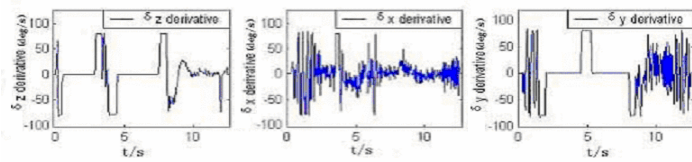


Fig. 3. The turn rates of elevator, aileron and rudder

In figure 3, although three general effectors don't exceed their upper limit($\pm 80^\circ/s$), they must deflect very quickly to follow the state commands for the flight states, however the phenomena is a very strict test for actuators which make them abrasion easily.

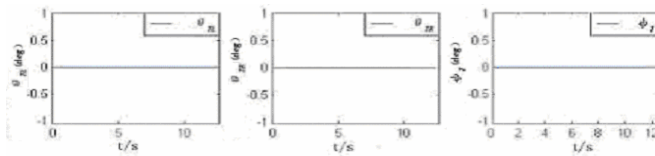


Fig. 4. The angle of twin thrust vector vanes

In figure 4, the left figure, middle figure and right figure respectively denote the deflection of twin thrust vector vanes $\theta_{TR}, \theta_{TL}, \varphi_T$. For the gain values corresponding to $\theta_{TR}, \theta_{TL}, \varphi_T$ being very larger up to 10000, there are no value assigned to θ_{TL}, θ_{TR} and φ_T , which assure that RHO controller can only assign signs to the general effectors.

From above analysis for figure 3, 4,5 and 6, it can be seen that thrust vector must be used to control fighter states to follow the state commands tightly within the extend flight envelop. Only using the general effectors are unreality to satisfy the relative objective.

Example 2. Simulation result of using the general effectors and thrust vector, and engines working at the max thrust state

Twin engines work at the max thrust state $30000N$. The positive gain matrices $Q(t)$ in this example is the same as the example 1. The general effectors and thrust vector are used within this chapter, but thrust value is const. The positive gain matrix $R(t)$ is set as: $R = \text{diag}(100, 5000, 5000, 15000, 2.5, 2, 0.5)$.

For keeping the const thrust value, the weighted value corresponding to P is always set as 100. Duo to the effect of thrust vector being considered, the gain values corresponding to θ_{TL} and θ_{TR} are all set to be 5000, and the one corresponding to φ_T is set to

be 15000, however these gain values are smaller than those values in Example 1., thus thrust vector have large active capacity. Three gain values corresponding to the general effectors are set as the same as values in Example 1.

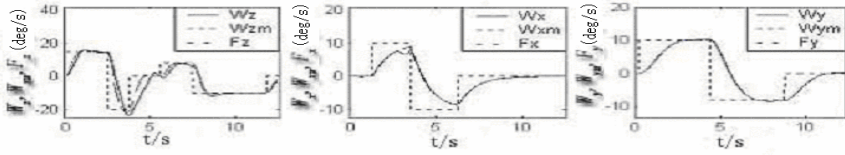


Fig. 5. The result of pilot command, flight state, and flight quality model

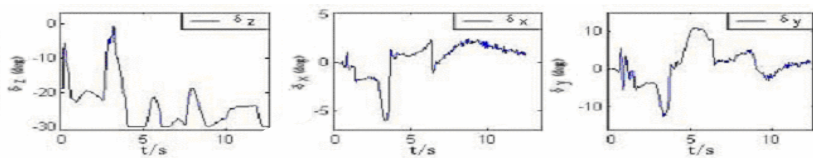


Fig. 6. The angle of elevator, aileron and rudder

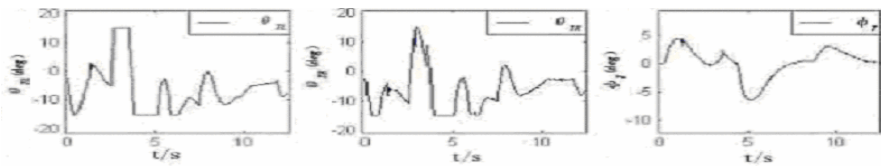


Fig. 7. The turn rates of elevator, aileron and rudder

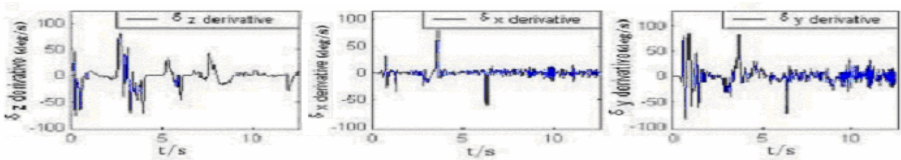


Fig. 8. The angle of twin thrust vector vanes

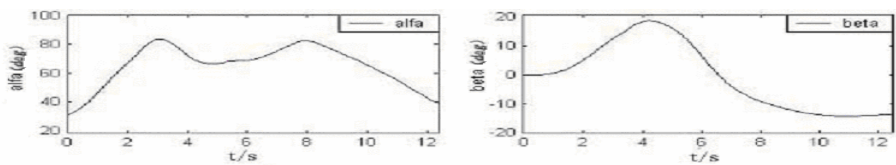


Fig. 9. The AOA and the the angle of sideslip

The descriptions of figure 5~8 are the same as figure 1~4. Figure 9 shows the mutative process of the AOA and the angle of sideslip.

Due to control signals assigned to thrust vector, the weight on the general effectors is alleviated, so three flight states can track the reference state commands tightly. But in figure 5 two directional following errors appear within the time section (3s, 7s) for that ω_z can't follow ω_{zm} (left of figure 5), within the time section (3s, 3.5s) for that ω_x can't follow ω_{xm} (middle of figure 5). In figure 6, it's seen that the general effectors are assigned smaller control signals (compare figure 6 with figure 2), which provide smaller force and moment of aerodynamic; however, thrust vector almost using all the control energy in three directions within figure 8, and both of θ_{TL} and θ_{TR} often reach their limit 15° , but thrust value can only keep 30000N, so thrust vector can't provide more pitch maneuver moment for ω_z to trace ω_{zm} tightly within the time section (3s,7s). Because θ_{TL} and θ_{TR} simultaneously affect longitudinal and lateral of flight, and in the optimal index matrix $Q(t)$, the gain value corresponding to errors of longitudinal is more important than that of lateral, which firstly assure following longitudinal command ω_{zm} at any time, thus when θ_{TL} and θ_{TR} get to the limit to follow longitudinal command ω_{zm} . They can't give attention to the lateral command ω_{xm} , which lead to within (3s,3.5s) that ω_x can't follow ω_{xm} .

The deflexion rates of the general effectors are smaller within figure 7 than those within figure 3, which mean burdens of actuators are alleviated.

In figure 11, the AOA (left figure) and the angle of sideslip (right figure) are shown. For producing larger lift aerodynamic, high AOA must be needed, at the same time, due to large variety of the state ω_z , the AOA vary very large, and keep 65° AOA for a long time, even up to 81° , so thrust vector must be used and large thrust value for aircraft to fly under so much larger AOA, which assure no flight quality is deteriorative.

From the above analysis, it can be seen that thrust vector can improve the maneuver capability of aircraft within the extend flight envelop greatly. However, due to the limit of the max thrust value, the states of flight can't follow the state commands at some time. Only increasing thrust value, the objective can be realized.

Example 3. Simulation result of using the general effectors and thrust vector, and engines working at the full afterburner state.

Twin engines work at the full afterburner state 40000N. The positive gain matrices $Q(t)$, $R(t)$ in this chapter are the same as the chapter 7.2. The constructs and descriptions of figure 10~14 are the same as figure 5~9.

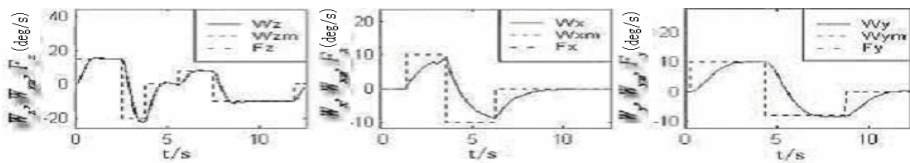


Fig. 10. The result of pilot command, flight state, and flight quality model

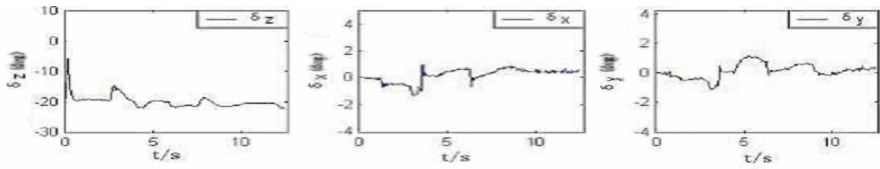


Fig. 11. The angle of elevator, aileron and rudder

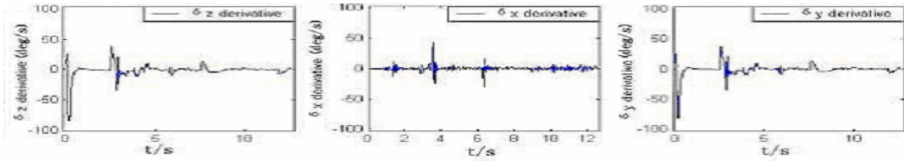


Fig. 12. The turn rates of elevator, aileron and rudder

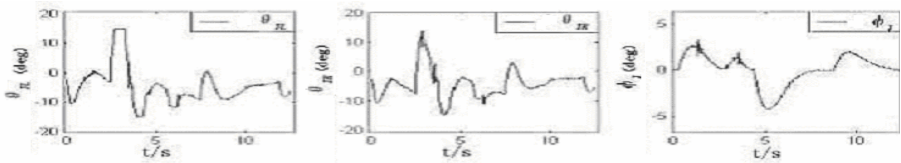


Fig. 13. The angle of twin thrust vector vanes

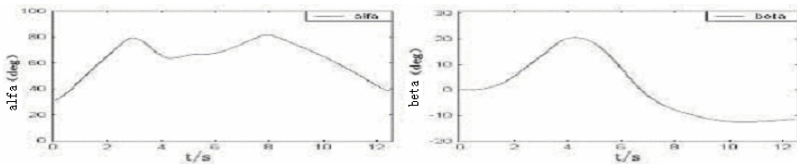


Fig. 14. The AOA and the the angle of sideslip

In figure 10, flight states can track the state commands exactly, due to thrust value increasing to the full afterburner state and thrust vector producing the needed moment of dynamic, thus the states in three directions are uncoupling entirely. In figure 11 and 12, the angle and the deflexion rates of the general effectors are smaller than those in figure 6 and 7. Due to increase of thrust value, thrust vector can produce the needed moment of aerodynamic and deflect smaller angle in figure 13 than those in figure 8.

5 Conclusion

In this paper, the test of manoeuvre efficiency within extend flight envelop is the objective. From the simulation result, we can find out how the thrust vector influence the maneuver efficiency within the extend flight envelop. First, the RHO control algorithm

is deduced. The simulation result show that at the max thrust state, aircraft without thrust vector can't track the reference command fairly well. Only using the general effectors within the extend flight envelop, it can achieve the requirement. By using effectors together with thrust vector, the maneuver capacity of aircraft improve greatly. However, the flight states often track the desired state commands inaccurately. In the third example, thrust values are set to the full afterburner and the states of three direction are fully uncoupling and track the state commands exactly. From the above simulations, we can see that the thrust vector and thrust value should work together and suitably, The aircraft can track the desire command within the extend flight envelop.

References

1. Su, H., Deng, J.: The Reconfigurable Flight Control Law Design for Aircraft with Thrust Vector. *Flight Dynamics* 20, 27–30 (2002) (in Chinese)
2. Javan, D., Roman, K.: Intelligent Adaptive Control of a Tailless Advanced Fighter Fighters (TAFA) in the Presence of Wing Damage. AIAA-99-4041 (1999)
3. Brinker, J., Wise, K.: Flight Testing of a Reconfigurable Control of a Tailless Fighter Fighters. AIAA-2000-3941 (2000)
4. Joseph, S., Kevin, A.: Reconfigurable Flight Control for a Tailless Advanced Fighter Fighters. AIAA-98-4107
5. Mayne, D.-Q., Michalska, H.: Rededing Horizon Control of Nonlinear System. *IEEE Trans. Automat. Contr.* 35, 814–824 (1990)
6. Parisini, T., Zoppoli, R.: A Receding Horizon Regulator for Nonlinear Systems and a Nueral Approximation. *Automatica* 31, 1442–1451 (1995)
7. James, W.: STOVL Integrated Flight Propulsion Control: Current Successes and Remaining Challenges. AIAA-2002-6021

Application of BP Neural Network in Stock Market Prediction

Bin Fang and Shoufeng Ma

School of Management, Tianjin University,
Tianjin 300380, China
Tckoguri@163.com

Abstract. Prediction for the change of stock market has been a hot research subject over the years. This thesis has introduced the definition and arithmetic of BP neural network model and established a stock market index prediction model based on the BP neural network model by taking advantage of the self-learning, self-adapting and nonlinear approximate ability. It is shown through empirical research that BP model not only has a rapid velocity of convergence and a high precision of prediction, but also has a certain application value if it is used for the short-term prediction of stock market index.

Keywords: Neural network, BP network model, Stock prediction, Relative error.

1 Introduction

Along with the rapid development of social economy, the stock market draws more and more attention of the public. Because the stock market is characterized by the coexistence of high income and high risk, the prediction for stock market index and stock price is always an issue that has always drawn people's attention. Two kinds of methods, namely fundamental analysis and technical analysis, are always taken in the prediction research of stock market. Fundamental analysis is a kind of macro analysis method, mainly studying such fundamental factors such as macro factor, industry factor and enterprise factor, which impact the trend of securities; technical analysis is a kind of micro analysis method, analyzing the number of transactions in the securities market and the trend of changes in prices with the technical means. The traditional technical analysis methods, like KD linear method, Dow Jones method and statistic method of regression analysis, are not ideal in the predictive effect for the stock market. As a kind of new space mapping method, the neural network has overcome the defects of traditional methods that are hard to identify the non-stationary state, but also can realize a complex causal relation and can draw off the nonlinear relation in the data aggregate automatically to conduct simulation, so it is a kind of most powerful tool for the nonlinear dynamic system prediction and model building and can better predict the short-term trend of the stock market index.

2 BP Neural Network Model

It can be divided into four typical structures according to the interconnected structure of neural network, namely feed forward network, feedback network, interconnected network and mixed network. Currently, many neural network models can be used for prediction, namely BP network, RBF network, genetic neural network and fuzzy neural network, etc. As BP neural network has self-adapting and self-organization ability and such features as strong generalization capability and strong fault tolerance capability, so this thesis chooses BP neural network to predict the stocks.

2.1 BP Neural Network

BP (Back propagation) model is a multi-layer feed forward network, adopting the learning approach of Minimum mean square error. It is a widely-used network and can be used in language integration, language identification, and self-adapting control, etc[1]. Its structure includes input layer, hidden layer and output layer. See Fig.1 for more details.

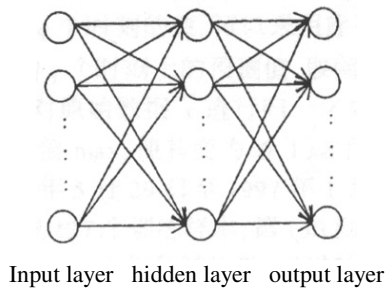


Fig. 1. Forward network

In BP model, the neurons between the layers realize an entire connection, which is to say that each neuron of the next layer realizes weight connection with each neuron of the previous layer and all neurons of each layer are not connected mutually and the connection weight of each layer of neurons in the network can be adjusted through learning. The learning process of BP neural network is divided into two stages: the first stage is forward-propagating, namely providing the input information and then forward propagating the input information onto the hidden nodes of the hidden layer through the input layer, and finally propagating the output information of hidden nodes into the nodes of output layer upon the operation of activation function of each unit and then obtaining the actual output value of each unit. The second stage is back-propagating: if the expected output value is not obtained from the output layer, the difference (namely error) between the actual output value and the expected output value shall be calculated and then the error signal returns along the original connection route; the error signal is propagated into the input layer to be calculated by modifying the weight of each layer of neurons and then reaches the allowable range through the continuous iteration of this forward-propagating process[2]. The self-learning calculation of the BP neural network model is an iterative procedure. By self-learning, BP network can give correct answers

not only for learned examples, but also for the models similar to the learned example, showing its strong ratiocinative ability which are suitable for solving nonlinear, large and complex problems.

2.2 Calculation Steps of BP Neural Network

There are many training calculation for neural networks, of which BP neural network is well-known for solid theory and wide application. The calculation steps are shown as follows[3]:

- ① Carry out initialization, give the random number to the weight matrix W and V and set sample mode counter p and training frequency counter q as 1, error E as 0 and learning rate η as the decimal value within the range of (0,1) and set the accuracy

E_{\min} after network training as a positive decimal value.

- ② Input the training sample and calculate the output of each layer, in which $y_j = f(V_j^T X) \quad j = 1, 2, \dots, m$ and $o_k = f(W_j^T Y) \quad k = 1, 2, \dots, l$.

- ③ Calculate the network output error:

$$E^p = \sqrt{\sum_{k=1}^l (d_k^p - o_k^p)^2}$$

- ④ Calculate the error signal of each layer:

$$\delta_k^o = (d_k - o_k)(1 - o_k)o_k \quad k = 1, 2, \dots, l$$

$$\delta_j^y = \left(\sum_{k=1}^l \delta_k^o w_{jk}\right)(1 - y_j)y_j \quad j = 1, 2, \dots, m$$

- ⑤ Adjust the weight of each layer:

$$\Delta \omega_{jk} = \eta \delta_k^o y_j = \eta (d_k - o_k) o_k (1 - o_k) y_j$$

$$\Delta u_{ij} = \eta \delta_j^y x_i = \eta \left(\sum_{k=1}^l \delta_k^o w_{jk}\right) y_j (1 - y_j) x_i$$

- ⑥ Check whether to complete one rotation training for all samples. If $p < P$, increase the counters p and q and then return to the Step②, or else return to the Step⑦;

- ⑦ Check whether the total network error reaches the accuracy requirement and set the total error as E_{RME} . If meeting $E_{RME} < E_{\min}$ the training finishes; or else setting E as 0 and p as 1, and return to the Step②.

It can be seen from the above steps that the weight shall be returned and weight shall be adjusted once a sample is inputted in the standard BP arithmetic, where the adjustment of weight is the core of arithmetic.

3 Stock Prediction Model Based on BP Neural Network

This model is a three-layer BP network structure with a hidden layer, where the selected data is the transaction data of Shanghai Stock Exchange Index from May 7 to September 9, 2008 (The data originates from the Dazhahui Software system).

3.1 Selection of Input Data

How to select the input data of BP neural network is a key issue, which directly impacts our classification results. Each component of input data shall select the quantitative indexes that can fully reflect the transactional features of stock market. The excess input data may complicate the data and reduce the network performance, while the fewer selected indexes are hard to make an accurate prediction. Through many times' test and analytical comparison, we have selected 12 common technical indexes in the stock market analysis. See Table 1 for more details.

Table 1. The components of input example vectors

λ_1	Open	λ_4	Low	λ_7	MA10	λ_{10}	K index
λ_2	Close	λ_5	Amount	λ_8	RSI	λ_{11}	D index
λ_3	High	λ_6	MA5	λ_9	BIAS	λ_{12}	J index

In the above table, it is easy to comprehend the components λ_1 — λ_7 and now we will make a simple introduction on the last five indexes. RSI refers to a kind of technical curve that is made according to the ratio sum of ascending and descending amplitude within certain period, which can reflect the booming degree within certain period. BIAS refers to the percentage of deviation and moving average line of stock index, which is an important supplement for the theory of moving average line. KD index is one of the technical analysis indexes, which integrates the advantages of relative strength index(RSI) and moving average line, has a sensitive reflection to the short and medium-term market situation, but also is a forceful tool for the technical analysis of short and medium-term stocks. When KD index is being introduced, it is always attached with a J index, where the calculation formula is $J=3D-2K=D+2(D-K)$, so the substance of J is to reflect the difference value between D and K.

3.2 Pretreatment of Input Data

The treatment of input data means that the data obtained from the stock market is converted into the data that can be identified by the neural network. As for Sigmoid function, its output is within 0 and 1, so it is necessary to conduct unitary processing for the sample data, and the most standard normalized treatment formula is: $\bar{x}_i = (x_i - x_{\min}) / (x_{\max} - x_{\min})$, in which x_{\max} and x_{\min} are respectively the maximum and the minimum in the sample data, x_i is the data of original

sample and \bar{x}_i is the converted value. As there is a great difference in the order of magnitude in each weight of stock market sample and the network training process will soon be controlled by the weight with a large order of magnitude, some data needs to be further treated[4]: $K=K/100$, $D=D/100$, $J=J/100$, $RSI=RSI/100$ and $BIAS=10 \cdot BIAS$; upon completion of treatment for neural network, a reversely normalized operation shall be made for those data again.

3.3 Confirming the Number of Neurons of Hidden Layer

The number of neurons of hidden layer is related to the number of neurons of input layer and output layer, but the specific quantitative relation is not established yet currently. If the number of hidden nodes is too small, it is hard for the network to obtain the information from the samples and it is insufficient to summarize and embody the rule of samples within the training set; if the number of hidden nodes is too large, the irregular content (like noise) in samples may be firmly remembered, thereby adding the network load and reducing the system efficiency and more seriously decreasing the generalization capability of network. The common formula used for confirming the hidden nodes is: $m = \sqrt{n+l} + \alpha$ or $m = \log_2^n$, in which m refers to the number of hidden nodes, n refers to the nodes of input layer, l refers to the number of output nodes and α refers to the constant between 1 and 10. This model separately conducts iterated training for many times for sample data under different hidden nodes, and then compares the mean value of relative errors obtained each time, selecting the number of nodes with minimum relative errors as the final number of hidden nodes. Through many times' comparison, the relative error is minimum when the training times are 10,000 and the number of hidden nodes is 6. Its drawing is shown in Fig. 2.

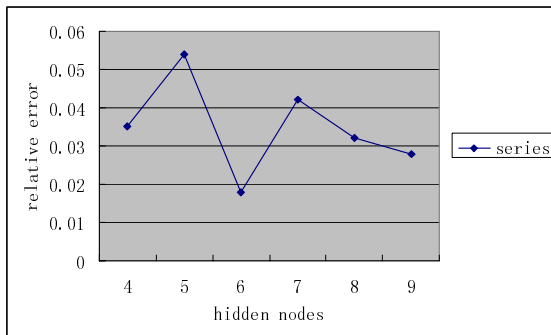


Fig. 2. Drawing of relative error and hidden nodes

3.4 Transfer Function

The effect of transfer function is to activate the neurons and then make them produce response to input, while proper transfer function can be selected in actual application. Sigmoid function is the most common activation function, mainly featured by nonlinearity, infinite differentiability and being similar to threshold function when the

weight is too large. The most common Sigmoid function can be divided into three kinds, namely $f(x) = 2/(1 + e^{-x}) - 1$, $f(x) = \tan hx$ and $f(x) = \arctan(x) \frac{\pi}{2}$. Through many times' test on the sample data of stock market, the results indicate that the function $f(x) = \tan hx$ is taken as the transfer function of hidden layer and output layer so as to obtain a better learning accuracy and a quicker convergence rate and minimum possibility of saturation.

3.5 Training and Prediction

The purpose of training is to find weight values with the smallest error within given number of iteration. After weight values are determined, predicting results may be gotten by inputting predicting example vectors. Through repeated test, the number of neurons of each layer in BP model is defined as 12-6-1, where the number of hidden

Table 2. The training results of the network's training

	Training value	Actual value	Absolute error	Relative error
Line Number:01	3693.5471	3755.6499	62.10281	0.0165
Line Number:02	3671.4940	3837.0759	165.5819	0.0432
Line Number:03	3688.4210	3791.5149	103.0939	0.0272
Line Number:04	3673.0873	3805.7480	132.6607	0.0349
Line Number:05	3662.6817	3735.7009	73.01924	0.0195
Line Number:06	3673.3047	3837.7410	164.4363	0.0428
Line Number:07	3684.1287	3816.5010	132.3722	0.0347
Line Number:08	3664.6291	3802.8511	138.2219	0.0363
Line Number:09	3637.6462	3782.3711	144.7249	0.0383
Line Number:10	3640.3721	3612.6389	27.73324	0.0077
Line Number:11	3617.5866	3718.9761	101.3894	0.0273
Line Number:12	3625.6878	3657.4939	31.80613	0.0087

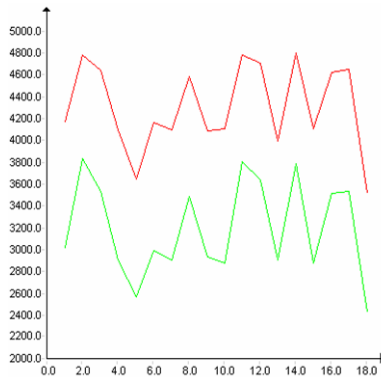


Fig. 3. The simulation drawing of predicted results

layers is 6, the learning rate $\eta=0.01$, factor of momentum $\alpha=0.6$ and shape parameter $\beta=0.45$. The first 80% data is taken to compose the training samples, while the last 20% data is taken to compose the inspection samples, thereby predicting the closing price of Shanghai Stock Exchange Index. See Table 2 for the final training results (The selected first 12 training values are shown in Table 2) and see Fig. 3 for the simulation drawing of predicted results.

It is known from Table 2 that the absolute error between the actual value and the training value is controlled under 170 points and the relative error is lower than 0.05. Figure 3 is the effect drawing incurred after the translation of matched curve of samples for prediction. It can be seen from this drawing that the predicted effect is quite ideal, which indicates that the model established in this thesis can better simulate the short-term trend of stock market.

4 Conclusion

It is shown by examples that the three-layer BP neural network model established in this thesis has such advantages as high prediction accuracy and quick convergence speed and the predicted results are satisfactory. If this model is applied for analyzing the stock data, it can provide some valuable objective information for the decision makers. If further improvements are made in terms of sample treatment, prediction methods and self-adapting capability of network, a better result will be obtained.

References

1. Zhong, L.: Artificial Neural Network and Convergent Application Technology, pp. 12–25. Science Press, Beijing (2007)
2. Jiang, J., Liang, Y.C.: Neural Network and its Application in Stock Market Prediction. Journal of Inner Mongolia University for Nationalities 10, 405–407 (2002)
3. Han, L.Q.: Theory, Design and Application of Artificial Neural Network, pp. 48–53. Chemical Industry Press (2005)
4. Zhang, Y., Wu, W.: Primary Investigation on Capturing the Stock Market Dark Horse with BP Neural Network. Operations Research and Management Science 13, 123–126 (2004)

A Research of Physical Activity's Influence on Heart Rate Using Feedforward Neural Network

Feng Xiao¹, Ming Yuchi¹, Jun Jo², Ming-yue Ding¹, and Wen-guang Hou¹

¹ School of Life Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China

² School of Information and Communication Technology, Griffith University, Queensland, Australia
m.yuchi@gmail.com

Abstract. Heart rate (HR) signal analysis is widely used in the medicine and medical research area. Physical activities (PA) are commonly recognized to greatly affect the changes of heart rate. However, the direct relationship between heart rate and physical activities is hard to describe. In this paper, a model using feedforward neural network with the function of HR prediction is designed. This model reflects the effect how PA affect HR. Experiments was conducted based on the reallife signals from a healthy male. The mean absolute error of the predicted heart rate was relatively small. The result shows the potential of the proposed method.

Keywords: Heart rate, Physical activities, Prediction, Neural network.

1 Introduction

As a noninvasive tool, Heart Rate (HR) signal analysis is widely used in the medicine and medical research area. It is recognized [1] that physical activities (PA) have great effects on the changes of heart rate. Currently, researches and applications that combine HR and PA signals mainly focused on: energy expenditure measurement [2], autonomic nervous system assessment [3] and sports research [4]. Few works have focused on how PA influenced HR: Pawar et al.[5] presented one body movement activity detection system which was based on ECG signal, but not HR. Meijer et al. [6] built a linear relationship between the HR and the body movements. However, the experiments were implemented in specific conditions and the body movement was recorded as the counted number of activities, which could not appropriately reflect the actual PA.

The main purpose of this paper is to build a prediction model using the feedforward neural network to reflect the effects of PA on the HR. The model was based on the author's previous work [1] where a predictor with two inputs (PA and HR) was designed. Similarly in this experiment, the subject was equipped with a portable HR and PA monitor, proceeded to perform normal daily activities without any special routine or restriction. Four synchronized time sequences were recorded: $HR(n)$ and $Acx(n)$, $Acy(n)$, $Acz(n)$, which are processed from the HR signal $hr(m)$ and three

acceleration signals $acx(l)$, $acy(l)$, $acz(l)$ respectively. In the previous work, the three acceleration signals were converted into one PA signal using a method of averaging. Here, $HR(n)$ and $Acx(n)$, $Acy(n)$, $Acz(n)$ are used as parallel inputs in the current time step to keep more original movement information., and the output is the predicted sequence $HR(n+1)$ in the next time step. Considering that all of the signals are non-constrained and real-time data, the predictor has the potential to be used in various areas, such as: cardiopathy research and diagnosis, heart attack warning indicator, sports capability measure and mental activity evaluation, etc.

As it is difficult to identify the direct rule behind the relationship between HR and PA, a feedforward neural network (FFNN) [7] was chosen as the mathematical model for the predictor for its intrinsic nonlinearity and computational simplicity. Levenberg-Marquardt algorithm was used in the training process of the FFNN.

2 The Research Method

2.1 HR Prediction Model

To investigate the relationship between the HR and PA, all the signals need to be recorded simultaneously. One portable HR and PA monitor from Alive Technologies was used here. The monitor measures and records the wearer’s ECG and PA (3-D acceleration) signals and determines the HR from the ECG in real-time. The left part of Fig.1 shows the subject (user) wearing the monitor. The specification of the monitor will be described in Section 2.2.

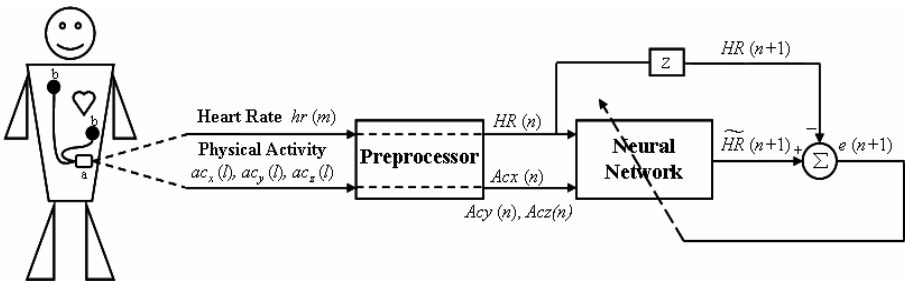


Fig. 1. The block diagram of the whole system

Table 1. Data Specification of Alive Heart Monitor

Signal	ECG	Accelerometer
Channels/Axis	Single Channel	3 Axes
Resolution	8 bits	8 bits
Sampling Rate	1 samples/sec	75 samples/sec
Dynamic Range	-2.66mV -2.66mV	-2.7g - 2.7g
Bandwidth	0.5Hz -90Hz	0Hz-20Hz

The middle part of Fig.1 is the preprocessor which converts the HR signal ($hr(m)$) and acceleration signals $acx(l)$, $acy(l)$, $acz(l)$ into usable format. The outputs of the preprocessor include four synchronized sequences: $HR(n)$ and $Acx(n)$, $Acy(n)$, $Acz(n)$, which are used in the FFNN as inputs. The output of the neural network is $HR(n + 1)$, which is the predicted HR in next time step.

2.2 Heart Rate and Physical Activity Recorder

Many studies on HR are based on the experimental data gathered in specific conditions and/or environments, whereas, this research was conducted with the data collected from normal daily activities, without any pre-planned routine. Consequently, a portable device is needed, which can monitor and record the HR and PA signals simultaneously for a period of time with relatively high accuracy. According to the device requirements, one commercial product Alive Heart Monitor (AHM) is chosen for our experiments. The collected data can be saved in an internal SD memory card or transmitted to PC, smart phone or PDA using Bluetooth in real time. The data specification of the AHM is shown in Table.1.

2.3 Signal Preprocess

The sampling rates of HR and acceleration are set differently in the AHM (1 samples/sec and 75 samples/sec, respectively) even though the inputs of the neural network are required to be sequences with same sampling rate. Here, $hr(m)$ and $acx(l)$, $acy(l)$, $acz(l)$ are converted into four synchronized sequences $HR(n)$ and $Acx(n)$, $Acy(n)$, $Acz(n)$ through a processing period τ .

Assume the whole recording period is T , the recorded data on each signal channel are evenly divided into N segments, each segment has the length of τ . When $\tau = 4s$, HR segment has 1 samples/s $\times 4s = 4$ samples (N_{hr}), and each acceleration segment has 75 samples/s $\times 4s = 300$ samples (N_{ac}). Then, the n th ($n = 1, \dots, N$) hr segment is converted (1) into $HR(n)$, and the n th acx , acy , acz segments are converted (2) into $Acx(n)$, $Acy(n)$, $Acz(n)$.

$HR(n)$ is the average (1) heart rate of n th segment. $Acx(n)$, $Acy(n)$, $Acz(n)$ are worked as average values (2) of the corresponding movements. However, instead of the HR signals being directly used, the absolute difference values of adjacent acceleration signals are adopted to calculate $Acx(n)$, $Acy(n)$, $Acz(n)$. This reflects the PA change between adjacent time steps.

It should be noted that the function of τ is not only to synchronize the inputs to neural network, but also to help to stabilize the prediction accuracy through averaging the noises. This works well, especially when some signals have high noises.

$$HR(n) = \frac{\sum_{m=(n-1)*N_{hr}+1}^{n*N_{hr}} hr(m)}{N_{hr}} \tag{1}$$

$$Acx(n) = \frac{\sum_{l=(n-1)*N_{ac}+1}^{n*N_{ac}-1} |acx(l+1) - acx(l)|}{N_{ac}} \tag{2}$$

2.4 Feed Forward Neural Network

In this work, there exist many factors which increase the difficulty of the prediction. The main factor is that the subject performs normal daily activities. The consequence is that the recorded HR is influenced by different aspects, such as, the subject's body condition, mood and surrounding environment.

These factors add uncertainties to the experiments. In fact, $HR(n)$ and $Acx(n)$, $Acy(n)$, $Acz(n)$, $HR(n)$ and $HR(n + 1)$ show nonlinear relationships in the data set obtained from the AHM, especially when τ is a relatively large value. Therefore, a mathematical method aiming at nonlinear prediction is needed. FFNN appears to be a good candidate [8]. With a certain structure, multi-layer FFNN can be used as a general function approximator [9].

Without needing any mathematical knowledge between the input and output, the FFNN [10] is trained based on comparisons of the output and the target, until the network reaches the goal.

Normally, the FFNN is trained with a backpropagation method, which includes many variations. Here, the Levenberg-Marquardt backpropagation algorithm [11] was adopted based on its steady performance of convergence and fast training speed for moderate-sized FFNN [12].

3 Experiment

3.1 Experiment Specifications

In this paper, the subject was a 33 years male with no record of heart disease. The recording time period was 12/90 minutes ($\tau=4/30s$). During this continuous period, the subject wore an AHM and performed the daily activities. To find the effect of the prediction interval τ to the predictor, the parameter τ was set to be $4s$ and $30s$ in two schemes. The recorded signals were evenly separated as two parts. In the two schemes, the first part of signals ($6min$ and $45min$ respectively) were adopted as the training set, which was used to train the FFNN; the remaining part of signals ($6min$ and $45min$ respectively) was for the test set, which was used to validate the trained neural network. Therefore, for both the training and test sets, $N = 90$. In Fig.2, we elucidate the corresponding $HR(n)$ and $Acx(n)$, $Acy(n)$, $Acz(n)$ of the training set and test set, which were preprocessed using (1) and (2), and will be used in our follow experiment.

C/C++ was chosen as the programming language. Two-layer FFNN was selected as the predictor for this experiment. The four inputs of the FFNN were $HR(n)$ and $Acx(n)$, $Acy(n)$, $Acz(n)$. The output layer (the last layer) had one neuron, $HR(n + 1)$, the predicted HR of the next time step. According to Kolmogorov theorem [13], the number of neurons in the hidden layer (first layer) we set is 9. Fig.3 shows the structure of the FFNN used in this paper.

The network was trained for 500 generations on the training set unless the train goal meets. Then it was tested on the test set.

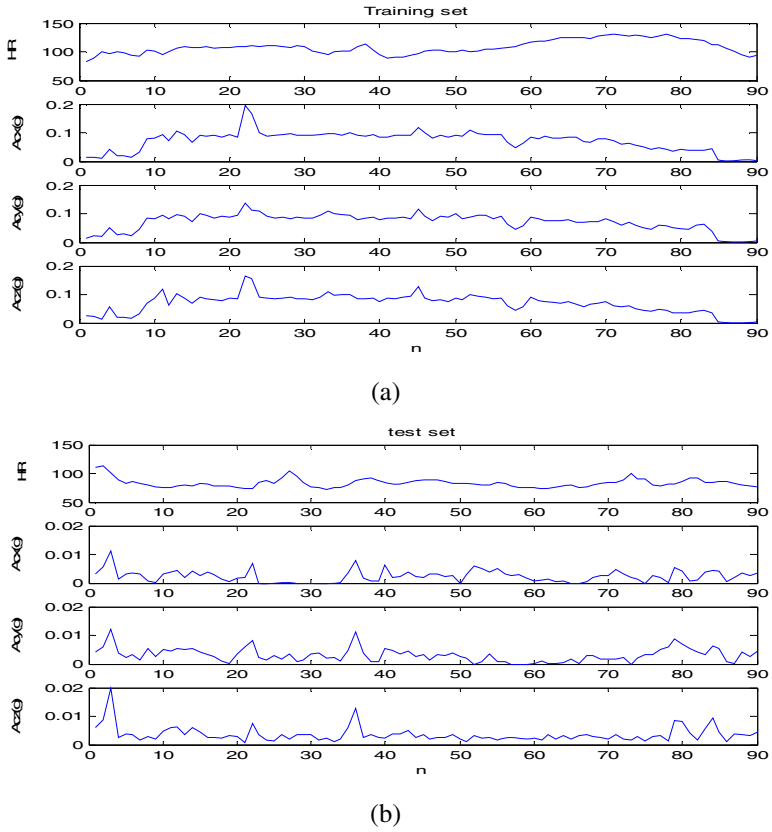


Fig. 2. Data sets for neural network training and validation, $T = 12\text{min}$, $\tau = 4\text{s}$, $N = 90$. $HR(n)/Acx(n)/Acy(n)/Acz(n)$ (a) Training Set; (b) Test set.

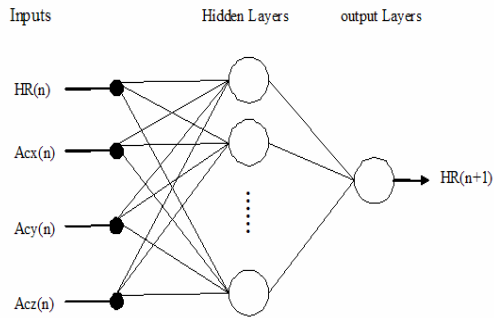


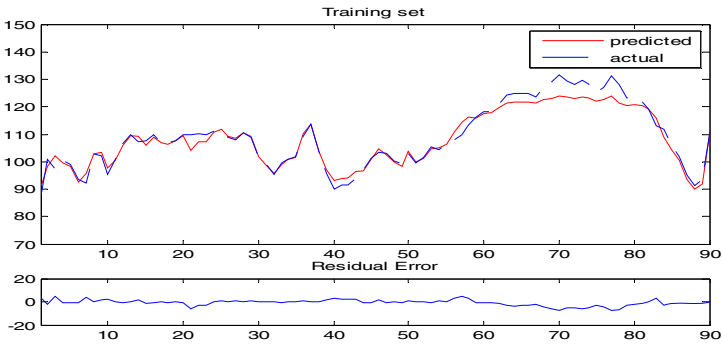
Fig. 3. Two-layer FFNN structure

3.2 Experiment Result

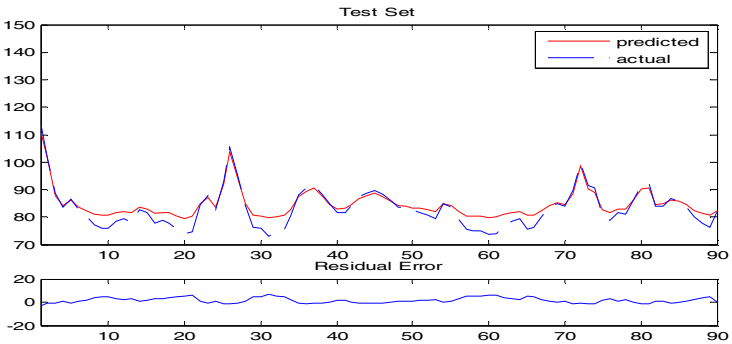
The performance of the neural network predictor on the training set and test set is shown in Fig.4. To make a clear identification, the predicted $HR(n + 1)$ is denoted with a red unbroken line, while the actual $HR(n+1)$ is represented by a blue dashed line. The figures indicate that the $HR(n + 1)$ follow the variance of $HR(n + 1)$ on both the training set and test set after training.

Table 2. The Mean Absolute Error between the actual and the predicted data

<i>MAEs(training/test)</i>	$\tau = 4s$	$\tau = 30s$
2 inputs	4.121/6.317	3.132/4.094
4 inputs	1.889/2.183	3.422/3.881



(a)



(b)

Fig. 4. Performance of the predictor($\tau = 4s$): predicted $HR(n+1)$ and actual $HR(n + 1)$, and the corresponding residual error. (a) Training set; (b) Test set.

The residual errors between the actual $HR(n + 1)$ and the predicted $HR(n + 1)$ are also shown in Fig.4. The corresponding mean absolute errors (MAEs) on training set and test set are 1.889 and 2.183 when $\tau = 4s$, 3.422 and 3.881 when $\tau = 30s$. Considering that the experiment was worked on real-life data, the MAEs on both training and test sets are acceptable. However, the variances of the error are still large relatively: 9.7236 and 12.0022 when $\tau = 4s$, 26.5524 and 24.8011 while $\tau = 30s$, respectively. It can be found that, some residual errors of test set are as big as 15 (when $\tau = 30s$), although most of the residual errors are smaller than 5. Table.2 compares the MAEs of 2 inputs and 4 inputs models. As it shows, the MAEs of 4 inputs model is smaller than that of 2 inputs model. But the results of 2 inputs model are steadier than that of the 4 inputs.

The MAEs of the predictor increased obviously as the prediction interval raised. However, a larger prediction interval could also bring more useful information into a same size training set and often represents a better performance of the predictor.

4 Conclusion and Discussion

In this experiment, prediction was performed every 4 seconds and 30 seconds. The result showed the potential of the predictor with the results close to the actual data. The mean absolute error could be restricted within a small range (inside 5). The consistency of the prediction needs be improved and will be addressed in the future work.

To validate the universal of the proposed method and improve the neural network performance, more and deeper investigations should be implemented. Firstly, more and various Data from subjects of varying age, gender and health level should be tested. Secondly, more tests on different system parameters, including prediction interval, total time length and sampling rate of the hardware. Thirdly, more Neural Network structures and types are needed. The RBF is another type of Neural Network which can be used as a predictor. And the most important factor to improve this system may be the PA Preprocess part. More useful signals preprocessed from PA (standard deviation, gradient) could be added into the predictor as inputs. The other possible varying factors include: data structure and training algorithm.

References

1. Yuchi, M., Jo, J.: Heart Rate Prediction Based on Physical Activity Using Feedforward Neural Network. In: International Conference & Hybrid Information Technology, Daejeon, Korea, August 2008, pp. 344–350 (2008)
2. Rennie, K., Rowsell, T., Jebb, S.A., Holburn, D., Wareham, N.J.: A Combined Heart Rate and Movement Sensor: Proof of Concept and Preliminary Testing Study. *European Journal of Clinical Nutrition* 54, 409–414 (2000)
3. Chan, H.L., Lin, M.A., Chao, P.K., Lin, C.H.: Correlates of the Shift in Heart Rate Variability with Postures and Walking by Time Frequency Analysis. *Computer Methods and Programs in Biomedicine* 86, 124–130 (2007)
4. Wang, W., Wei, J.R., Zhang, D.C., Xiao, S.Z., Wang, F.C.: A Study on 6-minute Walk Test Incorporating Cardiac Contractility and Heart Rate Change Measurements. *Chinese Medical Equipment Journal* 24, 16–18 (2003)

5. Pawar, T., Chaudhuri, S., Dutttagupta, S.P.: Body Movement Activity Recognition for Ambulatory Cardiac Monitoring. *IEEE Trans. on Biomed. Eng.* 54, 874–882 (2007)
6. Meijer, G.A., Websterp, K.R., Koper, H.: Assessment of Energy Expenditure by Recording Heart Rate and Body Acceleration. *Medicine and Science in Sports and Exercise* 21, 343–347 (1989)
7. Hagan, M.T., Demuth, H.B., Beale, M.H.: *Neural Network Design*. PWS Publishing, Boston (1996)
8. Hagan, M.T., Demuth, H.B., Beale, M.H.: *Neural Net and Traditional Classifiers*. Lincoln Laboratory, MIT. Tech., US (1987)
9. Hornik, K., Stinchcombe, M., White, H.: Multilayer Feedforward Networks are Universal Approximators. *Neural Networks* 2, 359–366 (1989)
10. Levenberg, K.: A Method for the Solution of Certain Non-linear Problems in Least Squares. *Quart. Appl. Math.* 2, 164–168 (1944)
11. Marquardt, D.: An Algorithm for Least-squares Estimation of Nonlinear Parameters. *SIAM J. Appl. Math.* 11, 431–441 (1963)
12. Hagan, M.T., Menhaj, M.: Training Feedforward Networks with the Marquardt Algorithm. *IEEE Trans. on Neural Networks* 5, 989–993 (1994)
13. Haykin, S.: *Neural Networks, A Comprehensive Foundation*. Macmillan College Publication, NewYork (1994)

Bi-directional Prediction between Weld Penetration and Processing Parameters in Electron Beam Welding Using Artificial Neural Networks

Xianfeng Shen, Wenrong Huang, Chao Xu, and Xingjun Wang

Institute of Machinery Manufacturing Technology, China Academy of Engineering Physics,
Mianyang 621900, China

Abstract. The bi-directional prediction between processing parameters and weld penetration benefits electron beam welding (EBW) production by reducing costly trials. An artificial neural network (ANN) model was established for the bi-directional prediction between them in EBW. The main processing parameters consist of accelerating voltage, beam current and welding speed, while weld penetration indicates penetration depth and penetration width of weld. The training and test sets were collected through EBW experiments by using 1Cr18Ni9Ti stainless steel. Two-layer supervised neural networks were used with different number of hidden layer nodes. Comparison between experimental and predicted results show the maximum absolute-value error is 6.6% in forward prediction from the main processing parameters to weld penetration, while that is 23.6% in backward prediction reversely. Combined the higher accurate forward prediction with the easy-use backward prediction in EBW production, a flow chart is proposed for optimizing prediction of processing parameters.

Keywords: Bi-directional prediction, Electron beam welding, BP neural network, Penetration depth, Penetration width, Processing parameters.

1 Introduction

Due to high-energy density, deep penetration, large depth-to-width ratio and small heat affected zone (HAZ) [1], electron beam welding (EBW) is widely used in aircraft, aerospace, machinery and other industries with requirement for low-distortion joints. Simultaneously, many EBW products are characterized by small quantities, various types of structure, and large differences in size. Therefore at presently, the procedure scheme and suitable processing parameters are determined by costly energy and time-consuming experimental trials for a new welding product. Forward prediction from the main processing parameters to weld penetration in EBW is achieved, namely, the weld penetration can be obtained with given processing parameters. Backward prediction from weld penetration to the main processing parameters is implemented, that is to say, a set of processing parameters can be got if you input the desired penetration. Consequently, bi-directional prediction between weld penetration and processing parameters is valuable to dramatically decrease the difficulty of determination of processing parameters, and to reduce the times of the costly experiments.

At present, predictions of weld penetration/shape and processing parameters in EBW are accomplished by three main ways: the finite element method based on the numerical simulation [2], statistical analysis [3, 4], and neural network [5]. Among them, the former two methods are trying to model the mechanism of welding processing, which require the thoroughly understanding the internal characteristics of welding process. Furthermore, the numerical simulation often needs to refine meshing for better accuracy, which greatly increases the complexity of the time-consuming calculations. Regression analysis, one of the statistical analyses, is a most commonly used methods in data-processing, but the analysis needs to determine the type of regression equation in advance which could become a very difficult task, especially with many input and output variables and complex coupling interactions.

Artificial neural network (ANN) does not need to fully understand the functional relation between input and output parameters, and can be trained iteratively from examples to learn and represent the complex relationships implied within the data. Moreover, the prediction with ANN is more accurate than that with regression equation. Predictions of welding performance and welding parameters based on ANN have been widely used in gas shielded arc welding, friction stir welding and other welding [6-9]. In this paper, an ANN model was established for the bi-directional prediction between the main processing parameters and weld penetration in EBW, and the training and test sets were collected through EBW experiments. Forward predictions from the main processing parameters to weld penetration and backward predictions reversely have been carried out by using the 2-layer ANN. A flow chart is proposed for optimizing prediction of processing parameters by combining the forward prediction with the backward prediction.

2 Problem Representation and Modeling

The typical cross-section shape of electron beam welding of 1Cr18Ni9Ti is shown in Fig.1, and looks like a 'nail' for deep penetrating of EBW. The main characteristic parameters of weld shape include penetration depth (H), penetration width (W). The former H refers to the distance between the weld root and specimen surface which is one of the most important parameters to evaluate the weld joints' effective weld depth, while the latter W denotes the weld seam width of melting surface. Some other important parameters of weld seam, such as aspect ratio (H / W), that is, penetration depth-to-width ratio, can be deduced from the two basic parameters H and W. Aspect ratio is also a very important parameter in welding production because with the greater aspect ratio as to the same penetration depth, the smaller the amount of heat input, the smaller welding deformation.

Accelerating voltage, beam current and welding speed constitutes the main electron beam welding processing parameters, and penetration depth, penetration width are important identification parameters of the weld shape. The forward prediction defined as the prediction from 'accelerating voltage, beam current, welding speed' to 'penetration depth, penetration width' while backward prediction refers to that from 'penetration depth, penetration width' to 'accelerating voltage, beam current, welding speed'. The bi-directional prediction used in the paper includes the forward prediction and the backward prediction mentioned above.

Feed-forward neural network and recurrent neural network are the most common ones among nearly 200-kind neural network models. The error back-propagation neural network (BP network) is commonly adopted in about 80% to 90% of all applications with feed-forward neural network. This article aims to establish the neural network model of the main processing parameters in EBW. The network should meet the following three conditions: continuous input, continuous output and supervised training, so the choosing of BP network is appropriate. In fact the general applicability of the technique has been demonstrated by the proof that BP networks with a single hidden layer, having sufficient number of neurons, using threshold or Sigmoid transfer function, are universal approximators. Therefore, a single hidden layer of BP neural network with hyperbolic Sigmoid transfer function is employed in this article. Forward and backward prediction models are respectively shown in Fig.2 (a) and (b).



Fig. 1. The typical cross-section shape of electron beam welding of 1Cr18Ni9Ti

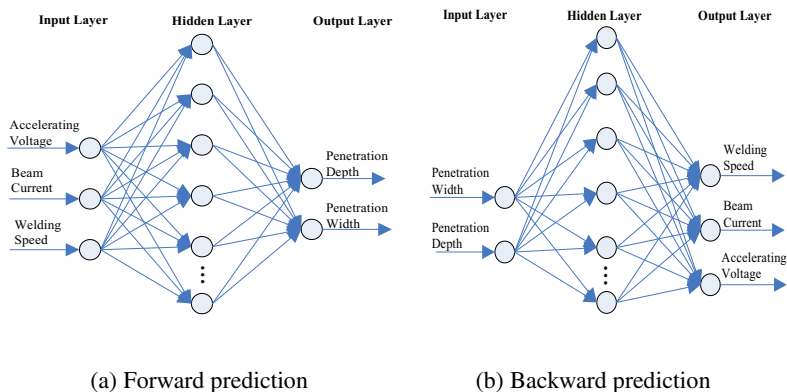


Fig. 2. The structure of three layered neural network of bi-directional prediction

3 Collecting of Experimental Data

All the training and testing samples were obtained through the test experiments. The welding machine in the experiments is high-pressure and high vacuum electron beam (EB) machine with its highest beam current of 100mA, maximum accelerating voltage of 150kV, a multi-purpose electron beam deflection generator, a vacuum chamber size of 1900mm × 1300mm × 1520mm, and maximum chamber vacuum degree of 5×10^{-4} mbar.

1Cr18Ni9Ti Cr-Ni austenitic stainless steel is taken as test material. It is known that austenitic stainless steels are prone to inter-granular corrosion and inter-granular stress corrosion cracking when they are subjected to sensitizing heat treatment between 723 and 1073 K, leading to premature failure of components during service [10]. Nevertheless, compared to carbon steel, austenitic stainless steel has a much larger expansion coefficient, leading to a larger welding deformation. However, during the electron beam welding of this material, the heat affected zone extends only to a narrow region across the weld pool, thus resulting in a lesser degree of defects and smaller deformation in the weld zone. Experiments were carried out by using surfacing welding for eliminating the effect of butt joint gap, and testing specimens with a dimension of 200 mm length, 160 mm width and 10 mm thickness.

A total of 29 kind EBW experiments were carried out by changing the processing parameters of acceleration voltage welding, beam current and welding speed with the experimental conditions and results as shown in table 1. The other parameters in welding were recorded as vacuum degree of not more than 5×10^{-3} mbar, deflection functions ∞ , deflection length and width of $0.2 \text{ mm} \times 0.3 \text{ mm}$, the deflection frequency of 1000Hz, and different focusing current for focusing on the sample surface. 26 were selected randomly from the 29 experimental results as neural network training sets, and the remaining 3 were taken as a test sets.

Table 1. Experimental conditions and results

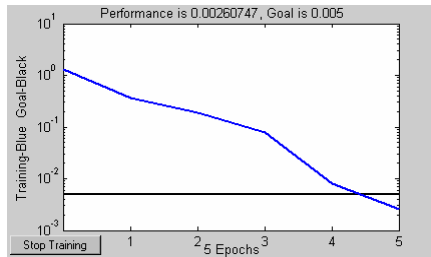
No.	<i>Factors</i>			<i>Experimental Results</i>	
	Accelerating voltage (kV)	Beam current (mA)	Welding speed (mm/s)	Penetration depth (mm)	Penetration width (mm)
1	120	5	13.3	2.35	1.62
2	120	10	13.3	5.33	1.63
3	120	19	13.3	10.00	1.60
4	120	5	20.0	1.98	1.19
5	120	25	20.0	10.00	1.43
6	120	25	26.7	8.75	1.26
7	100	5	10.0	1.93	1.93
8	100	10	10.0	5.19	1.99
9	100	20	10.0	10.00	2.20
10	100	20	20.0	10.00	1.80
11	100	5	15.0	1.67	1.55
12	100	25	15.0	10.00	1.99
13	100	30	15.0	10.00	1.71
14	80	5	10.0	0.91	1.42
15	80	10	10.0	2.04	2.16
16	80	25	10.0	5.06	3.56
17	80	30	10.0	6.27	3.87
18	80	35	10.0	7.41	4.15
19	120	16	13.3	8.79	1.67
20	120	20	20.0	8.99	1.28
21	100	20	15.0	9.01	1.70
22	80	20	10.0	3.90	3.21
23	120	13	13.3	6.88	1.65
24	120	15	20.0	6.60	1.29
25	100	15	15.0	6.52	1.75
26	100	15	10.0	8.33	2.08

4 Results and Discussion of Training and Testing of the Neural Network

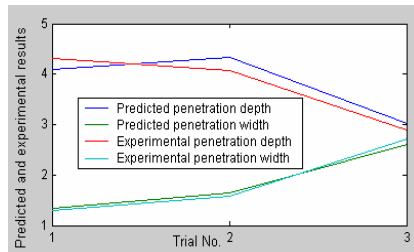
4.1 Training and Testing of the Forward Prediction Network

Training and testing of the forward prediction models have been undertaken with the use of training and testing set shown as in Table 1, and the other network training parameters are as follows:

- The neurons number of the hidden layer, $nNodeHide = 9$;
- Maximum number of epochs to train, $nEpochs = 2000$;
- Performance goal, $dbGoal = 0.005$;
- Epochs between showing progress, $nShow = 50$;
- The target error of testing set, $dbSset = 0.1$.



(a) Curve of network mean squared error vs. training epochs



(b) Comparison of the experimental and predicted results of testing data

Fig. 3. Training of the forward prediction network

Trainlm, a network training function, is employed in training of the network which updates weight and bias values according to Levenberg-Marquardt (LM) optimization. This LM algorithm appears to be the fastest method for training moderate-sized feed-forward neural networks (up to several hundred weights). Training process and results of the forward prediction network is shown in Fig.3, in which (a) represents curve of network mean squared error with training epochs, and it can be seen the network converges very quickly by the adoption of LM training function. From comparison between experimental and predicted results as shown in Table 2, it can be seen that percent error of penetration depth is range from -5.6% and 6.6% with

maximum absolute-value error 6.6%, and that of penetration width is range from -4.4% to 4.5% with maximum absolute-value error 4.5%. Therefore the network is precise enough to meet the requirement of predicting penetration depth and penetration width within a certain upper and lower limits in EBW production, which is helpful to decrease the testing times for determining appropriate processing parameters.

Table 2. Experimental data versus predicted results of the forward prediction

No.	<i>Experimental Results</i>		<i>Predicted Results</i>		<i>Percent error (%)</i>	
	Penetration depth (mm)	Penetration width (mm)	Penetration depth (mm)	Penetration width (mm)	Penetration depth	Penetration width
1	4.32	1.29	4.08	1.34	-5.6	3.9
2	4.07	1.57	4.34	1.64	6.6	4.5
3	2.90	2.72	3.03	2.60	4.5	-4.4
Maximum percent error of absolute value					6.6	4.5

4.2 Training and Testing of the Backward Prediction Network

The same training set and testing set were used in backward prediction as in forward prediction, and the difference lies in that the input data are exchanged with output one. The other network training parameters are used in backward prediction as follows:

- The neurons number of the hidden layer, nNodeHide = 13;
- Maximum number of epochs to train, nEpochs = 2000;
- Performance goal, dbGoal = 0.02;
- Epochs between showing progress, nShow = 50;
- The target error of testing set, dbSset = 0.4.

Table 3. Experimental data versus predicted results of the backward prediction

No.	<i>Experimental Results</i>			<i>Predicted Results</i>			<i>Percent error (%)</i>		
	AV (kV)	BC (mA)	WS (mm/s)	AV (kV)	BC (mA)	WS (mm/s)	AV (kV)	BC (mA)	WS (mm/s)
1	120	10	20	127.2	10.2	23.1	6.0	1.5	15.3
2	100	10	15	123.6	7.6	15.0	23.6	-23.6	0.3
3	80	15	10	74.8	17.3	11.2	-6.5	15.0	12.1
Maximum percent error of absolute value							23.6	23.6	15.3

Notes: AV - Accelerating voltage; BC - Beam current; WS - Welding speed.

Trainlm was still taken as the BP network training function. From comparison between experimental and predicted results as shown in Table 3, it can be seen that percent error of accelerating voltage is range from -6.5% to 23.6% with maximum absolute-value error 23.6%, that of beam current is range from -23.6% to 15% with maximum absolute-value error 23.6%, and that of penetration width is range from 0.3% to 15.3% with maximum absolute-value error 15.3%. The network has the ordinary accuracy of the predicted processing parameters which can be referred as the initial parameters by operators, and then be further adjusted for desired weld penetration.

The network in backward prediction converges much slower than that in forward prediction. As we can see in Fig.4 (a), curve of network mean squared error versus training epochs of backward prediction with 13 number hidden layer nodes, the network doesn't reach its convergence until it trains 62 epochs later, with target error taken 4 times as much as that of forward prediction.

The number of hidden layer nodes has a greater influence on network convergence in backward prediction. Fig.4 (a) ~ (d) show the curves of network mean squared error versus training epochs of backward prediction with different number of hidden layer nodes (nNodeHide). As presented in the figure, the network can not converge after 2000 epochs training with nNodeHide 7, and it converges more quickly with increasing nNodeHide. On the other hand, the network in forward prediction converges quickly after 4 epochs training. There exist great differences in convergence between forward prediction and backward prediction.

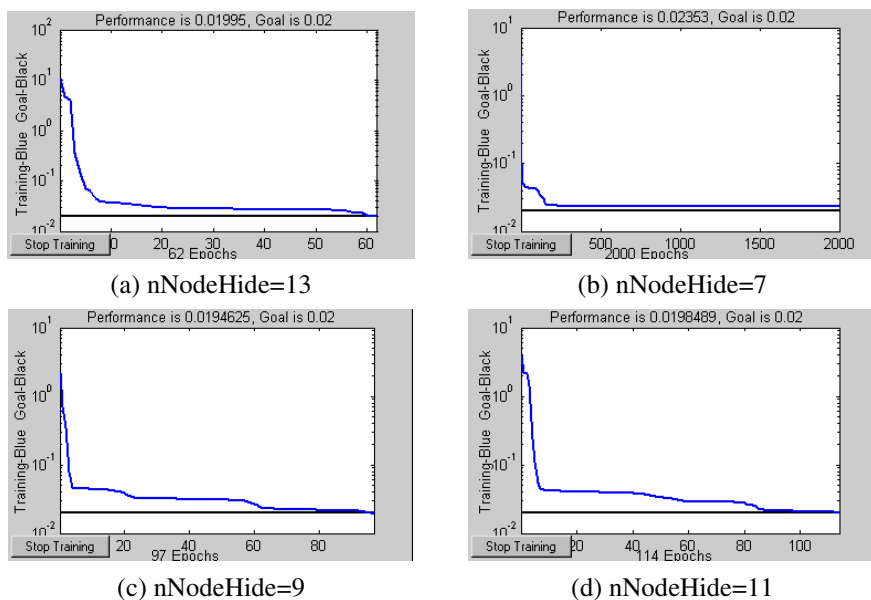


Fig. 4. Curves of network mean squared error vs. training epochs of backward prediction with different number of hidden layer nodes

4.3 The Flow Chart for Optimizing Prediction of Process Parameters

It can be seen from the comparison of training of forward prediction and that of backward prediction:

- (1) The network used in forward prediction is more accurate and faster convergence than in backward prediction. One possible reason is that the network in backward prediction is mapped from '2 dimension' to '3 dimension' which leads to more uncertain factors. Another likely reason is that coupling effect lies in processing parameters and several sets of parameters can lead to the same weld penetration in EBW.

(2) The backward prediction can be used more directly than the forward one in EBW production. As a matter of fact, operators in welding production would like to get a group of processing parameters from the network with input desired penetration depth and penetration width.

Is there any way to combine the advantage of more accurate forward prediction with the virtue of easy-use backward prediction? A flow chart for optimizing prediction of processing parameters as shown in Fig.5 was devised to attain this goal.

Several steps are involved in the flow chart. Firstly, input the target penetration depth and penetration width, and the initial parameters could be obtained by using the backward prediction. Secondly, round and adjust the initial parameters according to the experience, and check whether the goal penetration is reached with the use of the forward prediction. If the target penetration has not been attained, adjust the input processing parameters and go back the second step until the desired penetration is reached. At last, the satisfied processing parameters can be employed in the welding production. In this way, the forward prediction is united with the backward prediction.

The preliminary testing results show that the flow chart is useful to reduce trail times and lower production cost for it can provide welding production with more accurate processing parameters, and does not add much workload to operators.

5 Conclusion

Processing parameters and weld penetration in EBW has been predicted by using BP neural network, and the following conclusion can be reached.

The artificial neural network models for EBW is been established which provides the means of predicting the weld penetration and processing parameters. Comparisons between experimental and predicted results show that maximum absolute-value error in forward prediction from processing parameters to weld penetration is 6.6%, and that in backward prediction reversely is 23.6%.

Combined the higher accurate forward prediction with the backward prediction which can be used directly in EBW production, a flow chart is proposed for optimizing prediction of processing parameters which is valuable to be referred as processing parameters. The preliminary testing results show that the flow chart is useful to reduce trail times and lower production cost.

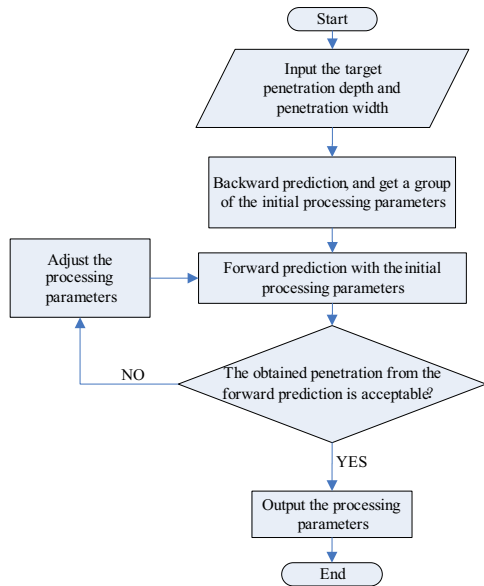


Fig. 5. Flow chart for optimizing prediction of processing parameters

References

1. Tian, Y.H., Wang, C.Q., Zhu, D.Y., Zhou, Y.: Finite Element Modeling of Electron Beam Welding of a Large Complex Al Alloy Structure by Parallel Computations. *Journal of Materials Processing Technology* 199, 41–48 (2008)
2. Wu, H.Q., Feng, J.C., He, J.S., Zhang, B.G.: Numerical Simulation of Deep Penetration in Electron Beam Welding of Ti3Al Intermetallic Compound. *Transactions of the China Welding Institution* 26, 1–4 (2005) (in Chinese)
3. Elena, K.: Statistical Modelling and Computer Programs for Optimization of the Electron Beam Welding of Stainless Steel. *Vacuum* 62, 151–157 (2001)
4. Koleva, E., Vuchkov, I.: Model-based Approach for Quality Improvement of Electron Beam Welding Applications in Mass Production. *Vacuum* 77, 423–428 (2005)
5. Kanti, M.K., Rao, S.P.: Prediction of Bead Geometry in Pulsed GMA Welding Using Back Propagation Neural Network. *Journal of Materials Processing Technology* 200, 300–305 (2008)
6. Bai, S.W., Tong, L.G., Liu, F.M., Wang, L.: Artificial Neural Network to Predict Toughness Parameter CVN of Welded Joint of High Strength Pipeline Steel. *Transactions of the China Welding Institution* 29, 106–108 (2008) (in Chinese)
7. Shi, Y., Li, J.J., Fan, D., Chen, J.H.: Predication of Properties of Welding Joints Based on Uniform Designed Neural Network. In: King, I., Wang, J., Chan, L.-W., Wang, D. (eds.) *ICONIP 2006*. LNCS, vol. 4234, pp. 572–580. Springer, Heidelberg (2006)
8. Hasan, O., Adem, K., Erol, A.: Artificial Neural Network Application to the Friction Stir Welding of Aluminum Plates. *Materials and Design* 28, 78–84 (2007)
9. Hakan, A.: Prediction of Gas Metal Arc Welding Parameters Based on Artificial Neural Networks. *Materials and Design* 28, 2015–2023 (2007)
10. Yilbas, B.S., Sami, M., Nickel, J., Coban, A., Said, S.A.M.: Introduction into the Electron Beam Welding of Austenitic 321-type Stainless Steel. *Journal of Materials Processing Technology* 82, 13–20 (1998)

Analysis of Nonlinear Dynamic Structure for the Shanghai Stock Exchange Index

Yu Dong and Hu Song

College of Science, Wuhan University of Science and Technology,
Wuhan 430065, China
yudong8502@126.com

Abstract. In this paper, we investigate the statistical properties of the Shanghai stock exchange index (SSEI). A GARCH-M(3,4) model and a TAR-ARCH-M(3,4) model successfully capture non-linear structure and asymmetries in the conditional mean and conditional variance. The TAR-ARCH-M(3,4) model is better in terms of forecasting performance.

Keywords: Nonlinear, Asymmetry, Conditional heteroscedastic, Stock exchange index.

1 Introduction

In the financial market research, people discovered that the change of asset prices disobeys the random walk model. What kind of model can describe the change of asset prices? Analysis of non-linear time series provided a tool for us to study volatility of the financial time series. The financial time series usually presents three following main characteristics: (1) heteroscedastic; (2) its distribution is asymmetry and heavy-tailed, as a result deviates Normal distribution; (3) Lever effect. The first characteristic was described in the autoregressive conditional heteroscedasticity model (ARCH) by Engle [1] or in general autoregressive conditional heteroscedasticity model (GARCH) by Bollerserlev [2]. The other two characteristics were described by TAR-ARCH model which was put forward by Zakoian [3] or EGARCH model by Nelson [4]. This paper is to get a model which can describe successfully non-linear structure of the Shanghai stock exchange index.

2 Basic Analysis of the Data

We collect data of daily closing stock price from the SSEI, the period of collection is from Jan. 2nd of 1997 to Dec. 30th of 2006 for 10 years.

Let P_t denote the daily closing stock price from the SSEI.

$$R_t = 100 \times \ln(P_t/P_{t-1})$$

R_t means daily income percentage. We compute some statistic of R_t , and result is shown in Table. 1.

Table 1. Statistics about R_t

Sequence	Mean	Standard deviation	Skewness	Kurtosis	Jarque-bear
R_t	0.0824	3.4919	5.1962	99.6014	789406.4

where Jarque-Bera is a value of test statistic JB. The JB is

$$JB = \frac{n}{6} \left(s^2 + \frac{(k - 3)^2}{4} \right)$$

where n is the sample size, s is the sample skewness, and k is the sample kurtosis.

The Jarque-Bera test is a two-sided goodness-of-fit test of the null hypothesis that the sample comes from a normal distribution with unknown mean and variance, against the alternative that it does not come from a normal distribution. Under null hypothesis the test statistic has a chi-square distribution with two degrees of freedom.

According to statistic, if the random variable sequences obey independent same normal distribution, its Skewness should equal zero and Kurtosis is three. If the random variable sequences obey normal distribution. Jarque-Bera statistic should obey a chi-square distribution with freedom 2, standard value of which is 9.21 and 5.99 under Significance Level being 1% and 5% respectively. From Tab.1, we find that the distribute of R_t far deviates normal distribute and it is heavy-tailed. The result of remarkable deviation suggests that R_t sequence may have non-linear dynamic structure.

3 Nonlinear Test

In order to examine whether the random variables are of independent identical distribution, Brock, Dechert and Scheinkman [5] developed an examination method, which is called BDS test.

The Brock proved $C_m(\varepsilon) = C_1(\varepsilon)^m$ with the hypothesis that random variables are independent identical distribution, and $BDS_m^n(\varepsilon)$ uniformly converge to standard normal distribution.

Here

$$x_t^m = (x_t, x_{t+1}, \dots, x_{t+m-1})$$

$$C_m(\varepsilon) = \lim_{n \rightarrow \infty} \frac{1}{(n - m)(n - m + 1)} \sum_{i \neq j} I(|x_t^m - x_j^m| < \varepsilon)$$

$$BDS_m^n(\varepsilon) = \sqrt{n} \frac{C_m^n(\varepsilon) - C_1^n(\varepsilon)^m}{\sigma_m^n(\varepsilon)}$$

where n denotes sample number, $I(\cdot)$ denotes indicator function and $\sigma_m^n(\varepsilon)$ is the estimate number of Standard deviation.

If the R_t is a random variables sequence of independent identical distribution by BDS test. the SSEI match to random walk model and explain that stock

market in Shanghai is weak and valid. Otherwise the sequence R_t exists a non-linear structure. How to choose a suitable non-linear model, we need to do a BDS test to model's standardize residual. If standardize residual sequence is tested to be a random variables sequence of independent identical distribution by BDS test, then the selected model is suitable. Otherwise we need to choose another non-linear model.

4 Nonlinear Model

BDS test is sensitive to deviate independent identical distribution, so it is very important to find the reason which refuse independent identical distribution. Whether the deviate come from the conditional heteroscedastic of sequence or distribute of asymmetry with fat tails? For answering this question, we will estimate three types of models: GARCH-M model, TARCH-M model and EGARCH-M model.

(1) GARCH-M model [6-8]

In equation of mean's estimate, we consider influence of variance or Standard deviation naturally, get GARCH-M model as follows:

$$R_t = c + \lambda h_t^2 + \varepsilon \text{ or } R_t = c + \lambda h_t + \varepsilon \tag{1}$$

$$\varepsilon_t | I_{t-1} \sim N(0, h_t^2)$$

I_t is a set of information at time t , and h_t^2 is conditional heteroscedastic (GARCH(p,q)), its form is as follows:

$$h_t^2 = \alpha_o + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j}^2$$

Among them, the p is the order of GARCH item and the q is the order of ARCH item.

(2) TARCH-M model

An important fault of GARCH model is to can't explain asymmetry, but the ARCH model which put forward by Zakoian(1994) can describe asymmetry. The equation of mean for ARCH model is the same as equation (1) in the GARCH-M model, and The equation about variance was changed as follows:

$$h_t^2 = \alpha_o + \sum_{j=1}^p \beta_j h_{t-j}^2 + \sum_{i=1}^q (\alpha_i \varepsilon_{t-j}^2 + \gamma_i \varepsilon_{t-j}^2 d_{t-j})$$

here

$$d_t = \begin{cases} 1 & \text{if } \varepsilon_t < 0, \\ 0 & \text{otherwise} \end{cases}$$

(3) The EGARCH-M model

The EGARCH-M model can describe asymmetry,

$$R_t = c + \lambda h_t^2 + \varepsilon \text{ or } R_t = c + \lambda h_t + \varepsilon$$

$$\varepsilon_t | I_{t-1} \sim N(0, h_t^2)$$

I_t is a set of information at time t , and h_t^2 is conditional heteroscedastic its form is as follows:

$$\ln(h_t^2) = \alpha_o + \sum_{j=1}^p \beta_j \ln(h_{t-j}^2) + \sum_{i=1}^q \left(\alpha_i \frac{|\varepsilon_{t-i}|}{h_{t-i}} + \gamma_i \frac{\varepsilon_{t-i}}{h_{t-i}} \right)$$

5 Model Analysis

5.1 The BDS Test of SSEI

We compute BDS statistics about R_t by using soft eviews, and result is shown in Tab. 2.

Table 2. The BDS test of SSEI

ε	m			
	2	3	4	5
0.01944	15.994	20.603	26.120	32.457
0.03888	16.707	20.391	23.060	24.947
0.07776	13.104	17.204	18.407	18.898

From result of BDS test, we should refuse the assumption that the R_t are random variables of independent identical distribution.

5.2 Nonlinear Model Fitting

By identification, we choose 3 order of R_t conditional variance for GARCH and 4 order for ARCH.

(1) GARCH-M(3,4) model fitting

$$R_t = -0.0547h_t + \varepsilon_t$$

$$\varepsilon_t | I_{t-1} \sim N(0, h_t^2)$$

$$h_1^2 = 0.6926 + 0.3644\varepsilon_{t-1}^2 + 0.3973\varepsilon_{t-2}^2 + 0.2098\varepsilon_{t-3}^2 + 0.4346\varepsilon_{t-4}^2 - 0.1900h_{t-1}^2 + 0.1069h_{t-2}^2 + 0.1896h_{t-3}^2$$

Next, we do a BDS test to model's standardize residual, and find the hypothesis that standardize residual is a sequence of independent identical distribution should be accepted under Significance Level is 5%. This expresses that GARCH-M(3,4) model fitting is successful.

(2) TAR-ARCH-M(3,4) model fitting

$$\begin{aligned}
 R_t &= -0.0362h_t + \varepsilon_t \\
 \varepsilon_t | I_{t-1} &\sim N(0, h_t^2) \\
 h_t^2 &= 0.6670 + 0.5672\varepsilon_{t-1}^2 + 0.3812\varepsilon_{t-2}^2 + 0.1879\varepsilon_{t-3}^2 + 0.3871\varepsilon_{t-4}^2 \\
 &\quad - 0.3607\varepsilon_{t-1}^2 d_{t-1} - 0.1857h_{t-1}^2 + 0.1128h_{t-2}^2 + 0.2077h_{t-3}^2
 \end{aligned}$$

Next, we do a BDS test to model's standardize residual, and find the hypothesis that standardize residual is a sequence of independent identical distribution should be accepted under Significance Level is 5%. This expresses that TAR-ARCH-M(3,4) model fitting is successful.

(3) EAR-ARCH-M(3,4) model fitting

$$R_t = -0.2103 + 0.1031h_t + \varepsilon_t$$

$$\varepsilon_t | I_{t-1} \sim N(0, h_t^2)$$

$$\begin{aligned}
 h_t^2 &= -0.551 - 0.984Ln(h_{t-1}^2) + 0.917Ln(h_{t-2}^2) + 0.921Ln(h_{t-3}^2) \\
 &\quad + 0.488 \frac{|\varepsilon_{t-1}|}{h_{t-1}} + 0.124 \frac{\varepsilon_{t-1}}{h_{t-1}} + 0.807 \frac{|\varepsilon_{t-2}|}{h_{t-2}} + 0.049 \frac{\varepsilon_{t-2}}{h_{t-2}} \\
 &\quad + 0.122 \frac{|\varepsilon_{t-3}|}{h_{t-3}} - 0.148 \frac{\varepsilon_{t-3}}{h_{t-3}} - 0.213 \frac{|\varepsilon_{t-4}|}{h_{t-4}} - 0.090 \frac{\varepsilon_{t-4}}{h_{t-4}}
 \end{aligned}$$

Next, we do a BDS test to model's standardize residual, and result is that we should refuse to accept the hypothesis that standardize residual is a sequence of independent identical distribution under Significance Level is 5%. This expresses that EAR-ARCH-M(3,4) model misfit.

5.3 The Model Evaluate

GARCH-M(3,4) model and the TAR-ARCH-M(3,4) can successfully describe nonlinear dynamic structure of the Shanghai stock exchange index, so which is better? In order to evaluate them, we need to compare their estimates effect. Therefore, we compute their root-mean-square precited error(RMSE):

$$RMSE = \sqrt{\frac{1}{100} \sum_{t=1909}^{2008} (R_t - \hat{R}_t)^2}$$

By the calculation, the RMSE of GARCH-M(3,4) model is 2.2539 and that of EAR-ARCH-M(3,4) model is 1.8576. Hence the TAR-ARCH-M(3,4) model is better in term of forecasting performance.

6 Conclusion

In this paper, we study the statistical properties of the Shanghai stock exchange index(SSEI). Our result shows three following points:

- (1) Currently stock market in Shanghai is not a mature market.
- (2) The wave of market is asymmetry and its distribution is heavy-tailed.
- (3) A GARCH-M (3,4) model and a TAR-ARCH-M (3,4) model successfully capture non-linear structure and asymmetries in the conditional mean and conditional variance. The TAR-ARCH-M (3,4) model is better in term of forecasting performance.

References

1. Engle, R.F.: Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom inflation. *Econometrica* 50, 987–1008 (1982)
2. Bollerslev, T.: Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics* 31, 307–327 (1986)
3. Zakoian, J.M.S.: Threshold Heteroskedastic models. *Journal of Dynamic Control* 18, 931–955 (1994)
4. Nelson, D.B.: Conditional Heteroskedasticity in Asset Returns. *Econometrica* 59, 347–370 (1991)
5. Brock, W.A., Dechert, W., Scheinkman, J.A., LeBaron, B.: A test for independence based on the correlation dimension. *Econometric Reviews* 15, 197–235 (1996)
6. Chou, R.F.: Volatility Persistence and Stock Valuations: Some Empirical Evidence Using GARCH. *Journal of Applied Econometrics* 3, 279–294 (1988)
7. Yi, D.H.: Data analysis and Eviews application. China statistics publisher 12, 186–200 (2003)
8. Pan, J.Z., Wu, G.X.: On tail behavior of nonlinear autoregressive functional conditional heteroscedastic model with heavy-tailed innovations. *Science in China (Series A)* 48, 1169–1181 (2005)

A Direct Approach to Achieving Maximum Power Conversion in Wind Power Generation Systems

Y.D. Song¹, X.H. Yin¹, Gary Lebby², and Liguo Weng²

¹ School of Electronics and Information Engineering, Beijing Jiaotong University, Beijing 100044, China

² North Carolina A & T State University, Greensboro, NC 27411, USA
National Institute of Aerospace, Hampton, VA, USA

Abstract. A new approach is proposed to achieve maximum wind power conversion by directly controlling wind turbine to operate along the maximum power coefficient curve (PCC). The control design is based on the so called “power coefficient dynamics”. It turns out that such dynamics are highly nonlinear and strongly coupled with uncertainties due to the involvement of both rotor dynamics and actuation (pitch) dynamics. Two set of control algorithms based on smooth variable structure control and memory-based control respectively are developed to ensure high precision PCC tracking, leading to high efficient power conversion. The effectiveness of the developed control algorithms are also verified via simulation.

Keywords: Wind power, Memory-based control, Pitch angle, Maximum energy conversion.

1 Introduction

Wind is a source of renewable sustainable power from air current flowing across the earth's surface. Wind turbines harvest this kinetic energy and convert it into usable electrical power. Showing in Figure 1 is the massive use of wind turbines in offshore [1]. Wind power represents the world's fastest growing energy source with an annual growing rate in excess of 30% and a foreseeable penetration equal to 12% of global electricity demand by 2020. Reliable and cost-effective electric energy generation from wind power relies on enabling technologies, including advanced control schemes. Various control methodologies have been reported in the literature on the subject of wind turbine controls [1]-[19].

Control of a DFIG wind turbine system was traditionally based on stator-oriented vector control [3-5], in which the control objectives were achieved with a rotor current controller. One main drawback of this system is that its performance depends highly on accurate machine parameters such as stator, rotor resistances, and inductances. Direct Torque Control (DTC) of induction machines provides an alternative to vector control [6]. This method directly controls machine torque and flux by selecting voltage vectors from a look-up-table using the stator flux and torque information. One

problem with the basic DTC scheme is that its performance deteriorates during starting and low-speed operations. Later on, Modified DTC approaches [7] [8] were developed. Other control methods include traditional PI based control [9], nonlinear and adaptive control [10] [11], yaw control [12-14], pitch control [15-18], and inverter firing angle control [19]. It is noted that most existing wind turbine control methods are essentially “indirect” methods in that they address the issue of wind power conversion efficiency indirectly through rotor speed adjustment to track the desired trajectory (which is designed based on maximum).



Fig. 1. An offshore wind farm

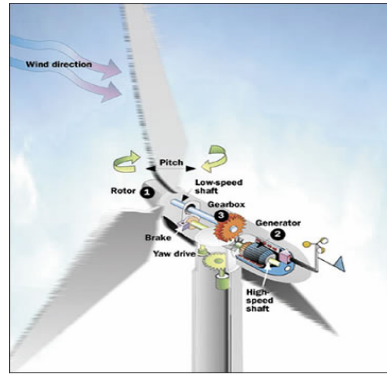


Fig. 2. Blade pitch control mechanism

In the paper, a direct approach is proposed to allow wind turbine to capture maximum power from available wind energy. The main idea behind this approach is to directly design a variable speed control strategy that forces the wind turbine to run along the maximum power coefficient curve (PCC) during its operation. This is achieved by adjusting blade pitch angle through the mechanism similar to the one as illustrated in Figure 2. Although several methods regarding pitch control have been reported [20-22], none of them addressed the power conversion issue directly. Furthermore, most results are based on the assumption that all parameters of the wind turbine are precisely known and no disturbance was acting on the system.

In this work, two sets of control algorithms are derived for maintaining high power conversion efficiency. The first one is Chattering-Free Variable Structure Control (CF-VSC), where the pitch angle rate (instead of pitch angle itself) is designed/specified as the control input, which, upon integration, leads to continuous and smooth real control input, removing the inherent chattering phenomena in traditional VSC. The second control algorithm is inspired by human memory/learning mechanism in which both rotor aerodynamics and pitch (actuation) dynamics are considered. The proposed memory-based control does not rely on precise system model, and demands much less computation as compared with most other methods. It learns from both past control experience and current observed system behavior to improve its performance. Both theoretical analysis and simulation studies confirm that the proposed method is able to achieve maximum power conversion efficiency.

2 System Description

2.1 Power Generation

The power produced by a given wind turbine can be determined by [23, 24].

$$P = \begin{cases} 0 & V < V_{in} \\ \frac{1}{2} C(\beta, \lambda) \rho A V^3 & V_{in} \leq V < V_R \\ P_R & V_R \leq V \leq V_{out} \\ 0 & V \geq V_{out} \end{cases} \quad (1)$$

where P is the generated power, A is swept area, V is the wind speed, V_{in} , V_R and V_{out} are the cut-in speed, nominal speed and cut-out speed, respectively, and $C(\beta, \lambda)$ is the power coefficient. Note that $C(\beta, \lambda)$ is a nonlinear function of the blade pitch angle β and the tip-speed-ratio (TSR) λ , as graphically shown in Figure 3 and Figure 4. The work in [25] attempted an analytic relation of $C(\beta, \lambda)$ versus β and λ , which is approximated by the following formula,

$$\left\{ \begin{aligned} C(\beta, \lambda) &= c_1 \left(\frac{c_2}{\lambda_i} - c_3 \beta - c_4 \right) e^{\frac{-c_5}{\lambda_i}} + c_6 \lambda \\ \frac{1}{\lambda_i} &= \frac{1}{\lambda + 0.08\beta} - \frac{0.035}{\beta^3 + 1} \\ c_1 &= 0.5176 \quad c_2 = 116 \quad c_3 = 0.4 \quad c_4 = 5 \quad c_5 = 21 \quad c_6 = 0.0068 \end{aligned} \right. \quad (2)$$

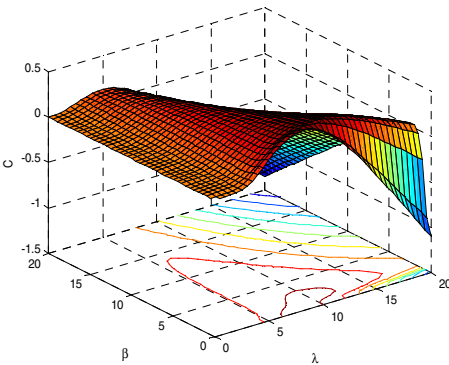


Fig. 3. A 3D view of PCC $C(\beta, \lambda)$

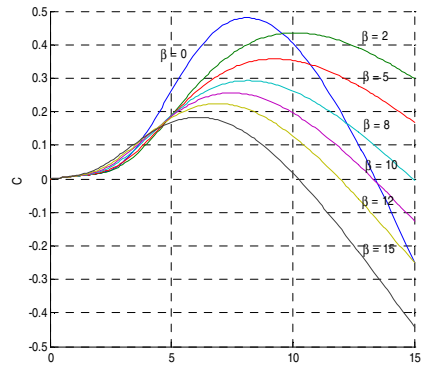


Fig. 4. PCC vs λ and β

2.2 Wind Turbine Rotor Dynamics

The dynamical model of a wind turbine can be described by [24] [26].

$$J \dot{\omega} = T_a - T_d(\cdot) \quad (3)$$

$$T_a = \frac{1}{2} C_r(\beta, \lambda) \rho ARV^2 \tag{4}$$

where J denotes the moment of inertia of the turbine-transmission-generator (all referred to the turbine shaft), T_a is the aerodynamic torque generated by wind flow, $C_r(\beta, \lambda)$ is a variable defined as $C(\beta, \lambda)/\lambda$, and $T_d(\cdot)$ is the resistant torque caused by friction, and electromagnetic damping, etc.

2.3 Pitch Dynamics

Wind turbine speed can be controlled by adjusting the blade pitch angle through the mechanism as shown in Figure 2. Most pitch based control methods ignore the pitch dynamics and assume that the pitch angle can be directly adjusted. A more practical method should take into pitch (actuation) dynamics into account. In this work, the following first order pitch dynamics is considered,

$$\dot{\beta} = \phi(\beta) + Ni \tag{5}$$

where i represents the driving motor current in the pitch mechanism, $\phi(\beta)$ is a nonlinear function and N is a constant. In next section, control algorithms based on pitch angle adjustment to achieve maximum power conversion efficiency are developed.

3 Control Design

3.1 Chattering Free-Variable Structure Control (CF-VSC)

A. Algorithm Description

First we define a control error $e = C - C^*$, where C^* is the desired power coefficient. Taking derivatives of both sides, yields,

$$\begin{aligned} \dot{e} &= \dot{C} - \dot{C}^* = \frac{\partial C}{\partial \beta} \dot{\beta} + \frac{\partial C}{\partial \lambda} \dot{\lambda} - \dot{C}^* \\ &= \frac{\partial C}{\partial \beta} \dot{\beta} + \frac{\partial C}{\partial \lambda} \left(\frac{R\dot{\omega}V - \dot{V}R\omega}{V^2} \right) - \dot{C}^* = \frac{\partial C}{\partial \beta} \dot{\beta} + \frac{\partial C}{\partial \lambda} \frac{R\dot{\omega}}{V} - \frac{\partial C}{\partial \lambda} \frac{\dot{V}R\omega}{V^2} - \dot{C}^* \end{aligned} \tag{6}$$

Note that: $J\dot{\omega} = T_a - T_d$, substituting $\dot{\omega}$ with $\frac{T_a - T_d}{J}$, gives,

$$\begin{aligned} \dot{e} &= \frac{\partial C}{\partial \beta} \dot{\beta} + \frac{\partial C}{\partial \lambda} \frac{R(\frac{1}{2\lambda} C \rho ARV^2 - T_d)}{JV} - \frac{\partial C}{\partial \lambda} \frac{\dot{V}R\omega}{V^2} - \dot{C}^* \\ &= \frac{\partial C}{\partial \lambda} \frac{RC\rho ARV}{2\lambda J} - \frac{\partial C}{\partial \lambda} \frac{\dot{V}R\omega}{V^2} - \dot{C}^* + \frac{\partial C}{\partial \beta} \dot{\beta} - \frac{\partial C}{\partial \lambda} \frac{RT_d}{JV} \end{aligned} \tag{7}$$

Rewriting the error equation as,

$$\dot{e} = f + g\dot{\beta} + \Delta f \tag{8}$$

where, $f = \frac{\partial C}{\partial \lambda} \frac{RC\rho ARV}{2\lambda J} - \frac{\partial C}{\partial \lambda} \frac{\dot{V}R\omega}{V^2} - \dot{C}^*$, $\frac{\partial C}{\partial \beta} = g$, $\Delta f = -\frac{\partial C}{\partial \lambda} \frac{RT_d}{JV}$. Unlike most existing methods in the literature, where the pitch angle β was directly designed, here the pitch angle variation rate is used as the virtual control input for developing the control scheme, this treatment leads to smooth VSC algorithms for the systems. To derive the control scheme, we need to assume that

$$|\Delta f| = \left| \frac{\partial C}{\partial \lambda} \frac{RT_d}{JV} \right| \leq \frac{R}{J} \left| \frac{\partial C}{\partial \lambda} \right| \left| \frac{1}{V} \right| |T_d| \leq c_d < \infty \tag{9}$$

where c_d is a constant. This assumption is reasonable and realistic because: 1) the external disturbance acting on the system is bounded; 2) R and J are system parameters, which are bounded; and the partial derivative of C with respect to λ is also bounded; and 3) V is wind speed, the inverse of which is a small number intuitively. The control scheme for pitch angle varying rate is generated by,

$$\dot{\beta} = g^{-1}(-f + u_c - ke) \quad \text{with } u_c = -\hat{c}_d \text{sign}(e) \tag{10}$$

where \hat{c}_d is the estimation of c_d , which is equal to $\int_0^t \gamma|e|$. Choose the Lyapunov function candidate

$$V = \frac{1}{2}e^2 + \frac{1}{2\gamma}(c_d - \hat{c}_d)^2 \tag{11}$$

with the proposed control in (10). It follows that

$$\begin{aligned} \dot{V} &= e\dot{e} - \frac{1}{\gamma}(c_d - \hat{c}_d)\dot{\hat{c}}_d = e(u_c + \Delta f - ke) - \frac{1}{\gamma}(c_d - \hat{c}_d)\dot{\hat{c}}_d \\ &\leq -|e|\hat{c}_d + |e|c_d - \frac{1}{\gamma}(c_d - \hat{c}_d)\dot{\hat{c}}_d - ke^2 \\ &= (c_d - \hat{c}_d)(|e| - \frac{1}{\gamma}\dot{\hat{c}}_d) - ke^2 \leq -ke^2 \end{aligned} \tag{12}$$

Therefore, the asymptotic stability is established, $t \rightarrow \infty, \hat{c}_d \rightarrow c_d$ and finally $|e| \rightarrow 0$. And moreover, it is seen that the pitch angle β is uniformly continuous by integrating.

B. Simulation Results

In this section, we simulated a scenario where the variation of wind speed was shown in Figure 6. The CF-VSC was applied to the wind turbine, and then the control performance was evaluated.

The parameters chosen for simulation are from [27]: $J = 32 \text{ kg.m}^2$, $R = 15 \text{ m}$, $\rho = 1.2 \text{ kg/m}^3$, $T_d = 2\sin(t) + 3\cos(2t)$. Once the wind speed and the system parameters are given, the corresponding optimal power coefficient curve as shown in Figure 6 can be acquired accordingly.

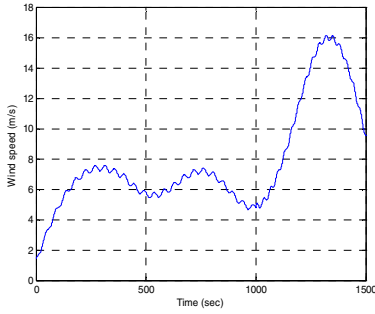


Fig. 5. Wind speed variation

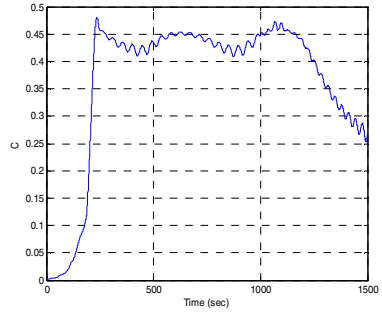


Fig. 6. Max power coefficient curve

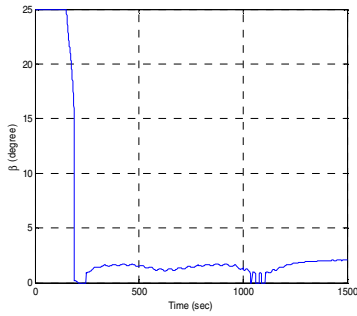


Fig. 7. Pitch angle variation

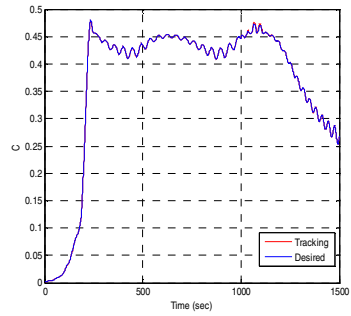


Fig. 8. Desired and actual power coefficients

The plot showing the adjustment of blade pitch angle β through the CF-VSC was given in Figure 7 and the desired versus actual power coefficients are plotted in Figure 8

The power generated by the wind turbine is shown in Figure 9. As seen from the results presented, the proposed controller works fairly well in maintaining maximum power conversion.

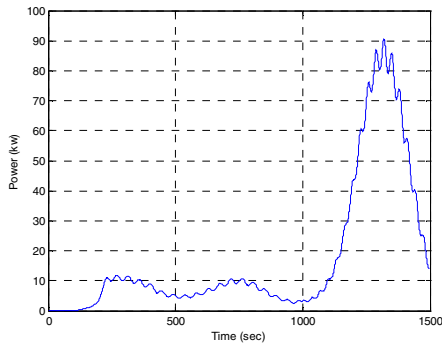


Fig. 9. Power output

3.2 Memory-Based Control

A. Algorithm Description

To apply memory-based control method to wind turbine systems, we recall the pitch actuation dynamics,

$$\dot{\beta} = \phi(\beta) + Ni \tag{13}$$

where i represents the driving current. Using both rotor dynamics and the actuation dynamics we have

$$\begin{aligned} \dot{e} &= \frac{\partial C}{\partial \lambda} \frac{RC\rho ARV}{2\lambda J} - \frac{\partial C}{\partial \lambda} \frac{\dot{V}R\omega}{V^2} - \dot{C}^* + \frac{\partial C}{\partial \beta} \phi + \frac{\partial C}{\partial \beta} Ni - \frac{\partial C}{\partial \lambda} \frac{RT_d}{JV} \\ &= Z + bi + \Delta \end{aligned} \tag{14}$$

where $Z = \frac{\partial C}{\partial \lambda} \frac{RC\rho ARV}{2\lambda J} - \frac{\partial C}{\partial \lambda} \frac{\dot{V}R\omega}{V^2} - \dot{C}^* + \frac{\partial C}{\partial \beta} \phi$, $b = \frac{\partial C}{\partial \beta} N$ is the system control gain, and $\Delta = -\frac{\partial C}{\partial \lambda} \frac{RT_d}{JV}$ represents the lumped uncertainties. The 1st order memory-based controller [28-30] is of the form,

$$\left\{ \begin{aligned} i &= (1 - \sigma(t))i_N + \sigma(t)i_A \\ i_{N,k} &= -ke \\ i_{N,A} &= w_0 i_{k-1} + b^{-1} \left(w_1 \frac{1}{T} e_k + w_2 \frac{1}{T} e_{k-1} + w_3 z_k + w_4 z_{k-1} \right) \end{aligned} \right. \tag{15}$$

$w_0 = 1, w_1 = -2, w_2 = 1, w_3 = -1, w_4 = 1$, which leads to

$$e_{k+1} = T(\Delta_k - \Delta_{k-1})$$

Therefore,

$$\|e_{k+1}\| \leq T^2 c_0 < \infty$$

where $c_0 = \max \left\| \frac{d\Delta}{dt} \right\|$ is the maximum possible variation rate of $\dot{A}(\cdot)$, which cannot be infinitely fast (otherwise no feasible control exists). Thus it can be concluded that $\|e\|$ is bounded, and PCC tracking stability is ensured.

B. Simulation Results

The simulation conditions and parameters used are the same as those for CF-VSC, except that T_d was replaced by $\sin(10t) + \cos(5t)$, which varies faster, therefore harder to deal with. The curve of control current was shown in Figure 10 and the tracking performance is illustrated in Figure 11.

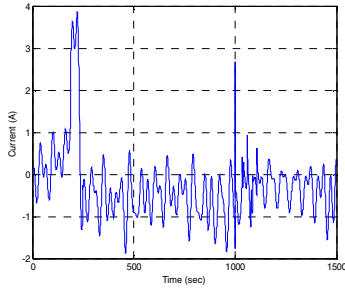


Fig. 10. Control current

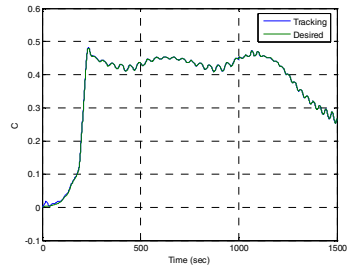


Fig. 11. Tracking performance

It is observed from the simulation that the memory-based control is able to achieve good PCC tracking with inexpensive on-line computations.

4 Conclusions

A new approach to achieve maximum power conversion in wind turbine systems is presented in this paper. Based on pitch angle adjustment through VSC and memory-based control algorithms, high power conversion efficiency is obtained by control the wind turbine to operate along the desired optimal PCC. The salient feature of the memory-based control lies in its simplicity and less dependence on detail system dynamics model. Yet only limited memorized system information is needed in building the control algorithms.

References

1. Freris, L.L.: Wind Energy Conversion Systems. Prentice-Hall, Englewood Cliffs (1990)
2. Muller, S., Deicke, M., Doncker, R.W.: Doubly fed induction generator systems for wind turbines. IEEE Industry Applications Magazine 8, 26–33 (2002)
3. Pena, R., Clare, J.C., Asher, G.M.: Doubly fed induction generator using back-to-back PWM converters and its application to variable-speed wind-energy generation. In: Proceedings of IEEE Electrical Power Application, vol. 143, pp. 231–241 (1996)
4. Akagi, H., Sato, H.: Control and performance of a doubly-fed induction machine intended for a flywheel energy storage system. IEEE Transactions on Power Electronics 17, 109–116 (2002)
5. De Doncker, R.W., Muller, S., Deicke, M.: Doubly fed induction generator systems for wind turbines. IEEE Industry Applications Magazine 8, 26–33 (2002)
6. Takahashi, I., Noguchi, T.: A new quick-response and high-efficiency control strategy of an induction motor. IEEE Transactions on Industry Applications 22, 820–827 (1986)
7. Habetler, T.G., Profumo, F., Pastorelli, M., Tolbert, L.M.: Direct torque control of induction machines using space vector modulation. IEEE Transactions on Industry Applications 28, 1045–1053 (1992)

8. Poddar, G., Joseph, A., Unnikrishnan, A.K.: Sensorless variable-speed controller for existing fixed-speed wind power generator with unity-power-factor operation. *IEEE Transactions on Industrial Electronics* 50, 1007–1015 (2003)
9. Ahmed, T., Noro, O., Matsuo, K., Shindo, Y., Nakaoka, M.: Wind turbine coupled three-phase self-excited induction generator voltage regulation scheme with static VAR compensator controlled by PI controller. *IEEE Transactions on Electrical Machines and Systems* 1, 293–296 (2003)
10. Dambrosio, L., Fortunato, B.: One step ahead adaptive control technique for a wind turbine-synchronous generator system. *IEEE Transactions on Energy Conversion Engineering Conference* 3, 1970–1975 (1997)
11. Sakamoto, R., Senjyu, T., Kinjo, T., Urasaki, N., Funabashi, T.: Output power leveling of wind turbine generator by pitch angle control using adaptive control method. *IEEE Transactions on Power System Technology* 1, 834–839 (2004)
12. Novak, P., Ekelund, T., Jovik, I., Schidtbauer, B.: Modeling and control of variable-speed wind-turbine drive-system dynamics. *IEEE Control System Magazine* 15, 28–38 (1995)
13. David, A.S.: *Wind Turbine Technology – Fundamental Concepts in Wind Turbine Engineering*. ASME Press, New York (1995)
14. Ekelund, T.: Yaw control for reduction of structure dynamic loads in wind turbines. *Journal of Wind Engineering and Industrial Aerodynamics* 85, 241–262 (2000)
15. Liebst, B.S.: Pitch Control System for Large-scale Wind Turbines. *Journal of Energy* 7, 182–192 (1982)
16. Johnson, G.L.: *Wind Power Systems*. Prentice-Hall Inc., Englewood Cliffs (1985)
17. Salle, S.A., Reardon, D., Leithhead, W.E., Grimble, M.J.: Review of wind turbine control. *Int. J. of Control* 52, 1295–1310 (1990)
18. Song, Y.D., Dhrikaran, B., Bao, X.: Variable Speed Control of Wind Turbines Using Nonlinear and Adaptive Algorithms. *Journal of Wind Engineering and Industrial Aerodynamics* 85, 293–308 (2000)
19. Hilloow, R.M., Sharaf, A.M.: A rule-based fuzzy logic controller for a PMW inverter in a stand alone wind energy conversion scheme. *IEEE Transactions on Industry Applications* 32, 57–65 (1996)
20. Senjyu, T., Sakamoto, R., Urasaki, N., Funabashi, T., Fujita, H., Sekine, H.: Output power leveling of wind turbine Generator for all operating regions by pitch angle control. *IEEE Transactions on Energy Conversion* 21, 467–475 (2006)
21. Kosaku, T., Sano, M., Nakatani, K.: Optimum pitch control for variable-pitch vertical-axis wind turbines by a single stage model on the momentum theory. In: *IEEE International Conference on Systems, Man and Cybernetics*, vol. 5, pp. 6–12 (2002)
22. Trudnowski, D., LeMieux, D.: Independent pitch control using rotor position feedback for wind-shear and gravity fatigue reduction in a wind turbine. *IEEE Proceedings on American Control Conference* 6, 4335–4340 (2002)
23. Mike, R., Veers, P.: *Wind Turbine Control Workshop*. Santa Clara University, Santa Clara (1997)
24. Abdin, E.S., Xu, W.: Control design and dynamic performance analysis of a wind turbine-induction generator unit. *IEEE Transaction on Energy Conversion* 15, 91–96 (2000)
25. Siegfried, H.: *Grid Integration of Wind Energy Conversion Systems*. John Wiley & Sons Ltd., Chichester (1998)
26. Wasynczuk, O., Man, D.T., Sullivan, J.P.: Dynamic behavior of a class of wind turbine generators during random wind fluctuations. *IEEE Transaction on PAS* 100, 2837–2845 (1981)

27. Chedid, R.B., Mrad, F., Basma, M.: Intelligent control of a class of wind energy conversion systems. *IEEE Transaction on Energy Conversion* 14, 1597–1604 (1999)
28. Song, Y.D.: Control of Wind Turbines Using Memory-based method. *Journal of Wind Engineering and Industrial Aerodynamics* 85, 263–275 (2000)
29. Song, Y.D.: Memory-based Control of Nonlinear Dynamic Systems Part I - Design and Analysis. In: 2006 1st IEEE Conference on Industrial Electronics and Applications, vol. 1, pp. 1–6 (2006)
30. Weng, L.G., Cai, W.C., Zhang, R., Song, Y.D.: Bio-Inspired Control Approach to Multiple Spacecraft Formation Flying. In: Second IEEE International Conference on e-Science and Grid Computing (e-Science 2006), p. 120. IEEE Computer Society Press, Los Alamitos (2006)

Synthetic Modeling and Policy Simulation of Regional Economic System: A Case Study

Zhi Yang, Wei Zeng, Hongtao Zhou, Lingru Cai, Guangyong Liu, and Qi Fei

Institute of Systems Engineering,
Huazhong University of Science and Technology,
Wuhan 43007, China
zengwei@mail.hust.edu.cn

Abstract. System dynamics and econometrics are proposed for a quantitative analysis of regional economy combined with qualitative analysis. The system dynamic model captures causal relationships of regional economic system, while econometrics is used to estimate equation parameters of system dynamics model. A system dynamics model for the development of regional economy is built by combination of system dynamics and econometrics. Finally, a case study of Shenzhen city gives some policy suggestions according to policy simulation results.

Keywords: System dynamics, Econometrics, Regional economy, Policy simulation.

1 Introduction

With China's reform and opening up, government's means of controlling macroeconomic gradually transformed into an indirect way. A variety of economic levers are used to adjust, affect macro-economic development situation to ensure sustained, stable, healthy and coordinated development of national economy. The impact of economic policies is not only limited to the economic field, but also brings synthetic influence to society and politics. Therefore, prospective study of economic policy is important.

How to analyze and enact economic policy is concerned by both the government and society. Because of the irreversibility of policy implementation and in order to cope with various contingencies of the world economy, it is necessary to perform simulation experiments on virtual economic system to analyze various possible consequences of economic policy.

2 Regional Economic System and Modeling

2.1 Regional Economic System (RES)

Regional Economic System (RES) is a subsystem of National Economic System (NES). RES is a high order and non-linear dynamic complex system. Its boundary is fuzzy and has multiple feedback loops. There are complex interdependent relationships among its subsystems [1].

The combination of system theory, economics and other scientific disciplines provide a good theoretical basis and technological means to study RES. For example, using dissipative theory and synergetic to study equilibrium, coordination and dynamic evolution of RES; using SD and gray prediction model to model RES[2,3]; traditional statistics models and econometric methods, such as DEA analysis of time series and input-output model, are also used to establish evaluation index system of RES[4,5]; With the development of computer science, artificial intelligence and distributed computer network have gradually been applied to the regional economic dynamic simulation system, early warning system and decision support system[6,7].

2.2 System Dynamics and Econometrics

System Dynamics (SD) is a quantitative method which is based on feedback control theory and computer simulation technology to analyze complex socio-economic system. It was founded in the middle 1950s by Professor Jay W. Forrester at MIT. Since SD takes large-scale systems, non-linearity as well as person's decision-making role into account, it is useful to solve the cyclical and long-term problems such as policy simulation. Econometrics is derived from economics, mathematics and economic statistics. Econometrics is used to build mathematical models of economic system according to economic theory and statistics. Econometric models are used to predict economic trend and make economic policy.

In this paper, SD and econometrics are combined to model regional economic system. In details, the basic model framework of RES is built using SD, while econometrics is used to analyze and estimate equation parameters of system dynamics model.

It brings several advantages to combine econometrics with system dynamics. First, it improves the accuracy of SD model' parameters. Then, it increases the reliability of mid-short term prediction. Last, it overcomes the disadvantage of econometrics which is difficult to analyze complex nonlinear systems and highly depends on the statistical data.

3 Regional Economic Development Forecasting Model and Policy Analysis about the City of Shenzhen

In the paper, synthetic modeling and policy simulation of regional economic system is illustrated with Shenzhen city. The main reasons are listed as follows: Shenzhen is a typical city of immigration with an obvious feature of large amount of floating population. Nearly a decade of high-speed economic development of Shenzhen brings increasing pressure of bearing capacity of environment and resources. Regulating investment structure is an important means to optimize industrial structure and employee's structure, and so guarantee a sustained, steady, coordinated growth of national economy.

3.1 Sustainable Development and Population Movement of City

The whole system's main framework includes 5 sub-systems: population sub-system, industry sub-system, investment sub-system, science, technology and education sub-system and resources sub-system. The main relationships among the five subsystems are shown in Fig. 1.

According to our research purpose, the population sub-system, industry sub-system and investment sub-system are mainly considered in our study, without ignoring the impact of science, technology and education sub-system on industry. The constraints of resources sub-system on the population and industry sub-systems are also considered in the paper.

The influence of resources sub-system on the population and industry is regarded as an environment influence factor. The influence of science, technology and education sub-system on the industry sub-system is treated as a technical influence factor in Cobb-Douglas production function. So our basic model not only highlights the relationships among the main sub-systems, but also considers the environmental and scientific influences on the economic development.

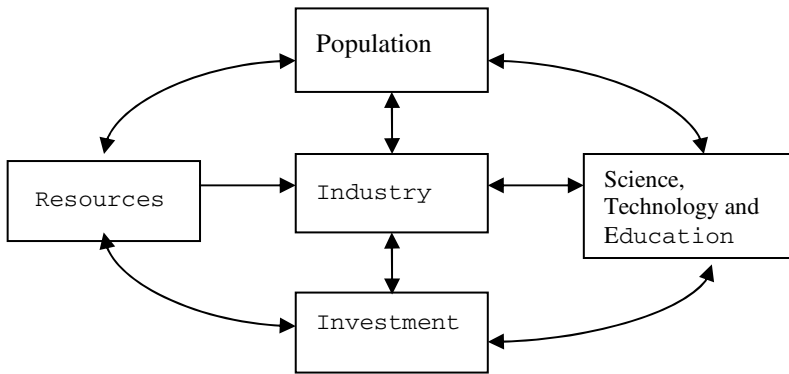


Fig. 1. Regional economic system framework

3.2 Economic Development of Shenzhen

Shenzhen as a typical immigrant urban, the proportion of the transient population to the resident population is above 70 percent since 1990s. After reform and opening up, the proportion of primary industry investment to the state-owned fixed assets investment (mainly refers to the infrastructure and the investment for renovation and transformation) in Shenzhen has dropped from 50% to less than 1%. The proportion of secondary industry investment and tertiary industry investment to state-owned fixed assets investment maintains respectively about 50%. From the viewpoint of gross domestic product (GDP), the three major industries keep rapid growth. From 1980 to 2005, the average growth rate per annum of primary industry is 12.1 percent; the average growth rate per annum of secondary and tertiary industry is about 43% and 37.4 %.

The average growth rate per annum of per capita GDP is 20.3%. Although the industrial structure and economic environment of Shenzhen is unique, it is proven that a reasonable industrial structure can bring a high-speed economic development.

3.3 Regional Economic System Dynamics Model of Shenzhen

3.3.1 Douglas Function

In economics, the Cobb-Douglas functional form of production functions is widely used to represent the relationship of an output to inputs. Its basic form is:

$$Y = A(t)L^\alpha K^\beta \mu \quad (1)$$

Where:

Y = total production (the monetary value of all goods produced in a year)

L = labor input

K = capital input

A = total factor productivity

α and β are output elasticities of labor and capital, respectively. These values are constants determined by available technology.

μ expresses the influence produced by random disturbance, $\mu \leq 1$.

The production function points out that the main factors determining the level of industrial development are labor force, fixed assets and integrated skills (including managerial and administrative expertise, labor force quality, technology and so on).

Formula 1 is easily transformed to formula 2 in order to facilitate data processing, where we assume that $B(t) = \mu A(t)$

$$\ln Y = \ln B(t) + \alpha \ln L + \beta \ln K \quad (2)$$

3.3.2 Primary Industry

According to statistical data of Shenzhen, the proportion of employment, investment and GDP output value of primary industry to three industries are very small. So in the paper, we will mainly consider the second and tertiary industry in the economic system dynamics model.

3.3.3 SD Model

The data used to establish SD model's equations and their main parameters are based on the statistical data released by Shenzhen municipal statistical bureau [8]. In the system dynamic model, econometrics is used to estimate model parameters from statistical data.

Fig. 2 shows the basic economic development system dynamics model of Shenzhen, which describes the main relationships among investment, industry and population sub-system. There are six state variables, six rate variables, thirteen auxiliary variables and some constants in the model. $\ln(*)$ denotes logarithm operation on parameter *.

3.3.4 Models Verification

In Table 1, compared the simulation data produced by the SD model to the statistical data between 1995 and 2005, almost all the errors are under 5%. As the simulation data of population is very close to the statistical data, we consider the SD model is valid.

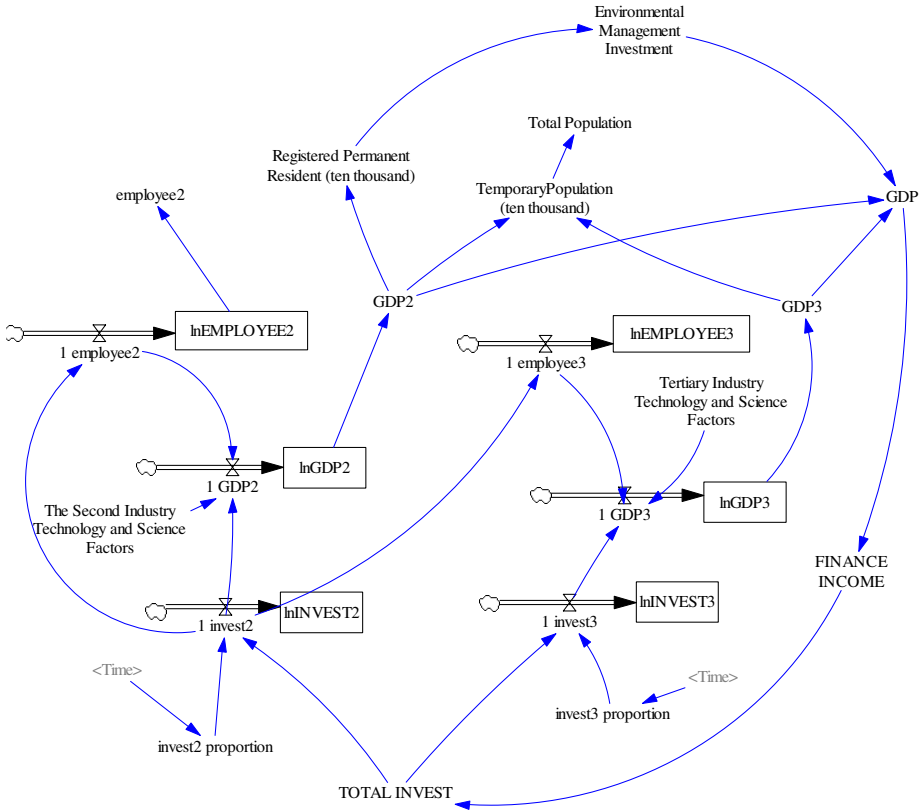


Fig. 2. SD model of Shenzhen regional economic system

The SD model tries to reflect and quantify the internal relationships among investment structure, GDP, industrial structure and population structure of Shenzhen. According to simulation results the SD model is accurately reflects the internal relationships among investment, industrial development and population structure.

Therefore, the SD model can be used to simulate the policy’s impact on fixed assets investment, GDP growth and population structure of Shenzhen, and to provide decision support for government.

3.4 Model Simulation and Policy Analysis

3.4.1 Forecast and Simulation from 2006 to 2015

Based on the simulated data of 2005 and keeping influencing factors of technology and environment invariability, we set the ratio between the secondary and tertiary industry's fixed assets investment and simulate from 2006 to 2015 respectively to forecast the total population, GDP and per capita GDP. The ratio sets to 5:5 (referred to as 55-mode), 4:6 (referred to as 46-mode), 3:7 (referred to as 37-mode), just as shown in table 2.

Table 1. Simulation results (1995-2005)

Year	Simulation total population (ten thousand)	Statistical total population (ten thousand)	Simulation GDP (hundred million Yuan)	Statistical GDP (hundred million Yuan)	Simulation per capital GDP (Yuan)	Statistical per capital GDP (Yuan)
1995	450.078	449.15	825.692	842.4833	18291.61	19550
1996	481.803	482.89	990.761	1048.442	20513.46	22498
1997	516.487	527.75	1186.57	1297.421	22927.25	25675
1998	553.867	580.33	1419.15	1534.727	25579.36	27701
1999	593.343	632.56	1699.67	1804.018	28605.66	29747
2000	635.004	701.24	2030.03	2187.452	31931.88	32800
2001	678.446	724.57	2420.47	2482.487	35642.38	34822
2002	723.142	746.62	2883.29	2969.518	39840.07	40369
2003	769.032	778.27	3429.82	3585.724	48341.34	47029
2004	815.561	800.8	4078.35	4282.143	49979.65	54236
2005	862.872	827.75	4845.22	4950.908	56127.42	60801

Table 2. Simulation data in three investment modes

Year	Simulation total population of Shenzhen (ten thousand)			Simulation GDP of Shenzhen (hundred million Yuan)			Simulation per capital GDP of Shenzhen(Yuan)		
	5:5	4:6	3:7	5:5	4:6	3:7	5:5	4:6	3:7
2006	910	910	910	5753	5753	5753	63155	63155	63155
2007	959	959	959	6828	6829	6834	71134	71162	71207
2008	1010	1010	1010	8101	8105	8114	80196	80248	80338
2009	1062	1062	1062	9609	9615	9630	90441	90515	90648
2010	1117	1117	1117	11394	11403	11425	101960	102053	102227
2011	1176	1176	1176	13507	13520	13550	114824	114931	115145
2012	1240	1240	1241	16009	16026	16066	129078	129195	129444
2013	1310	1311	1312	18972	18993	19045	144735	144858	145137
2014	1389	1390	1391	22479	22505	22571	161762	161885	162189
2015	1478	1479	1481	26630	26664	26746	180080	180196	180516

3.4.2 Simulation Results Analysis

From the viewpoint of economic development, 37-mode is one investment mode to get a high-speed economic development. The tertiary industry level is one of the most important modernization indices of a country or region. The development of tertiary industry is an important way to raise the living standards and to promote the market economy growth. Tertiary industry development must be based on the development of the first, second industry. In its initial development stage, it can attract the massive transient population and solve the partial population employment problem. But, it brings the great population and resources pressure to Shenzhen's environment at the same time.

From the viewpoint of population, 55-mode is a valid investment mode to avoid large amount of transient population emerging. But, in this mode, per capita GDP is lowest in three investment modes.

From the comprehensive viewpoint, 46-mode may be a more reasonable compromise investment mode according to the development situation of Shenzhen, whose industry investment must maintain at a reasonable level and introduce advanced manufacturing industries of energy saving, high output, low pollution and technology advanced. Only this can help Shenzhen to adjust and promote its manufacturing industries level and enhance the level of synthesis management of the second industrial. At the same time, environment excessive consumption often has the latency and delay characteristics, so the essential environmental resources investment is needed.

4 Conclusions

The system dynamics (SD) is mainly used to analyze and study non-linear information feedback system and has already been widely applied in many domains such as sociology, economics, and management science and so on. The SD model is regarded as a valid policy simulation tool and can show its advantage in the analysis of non-linear complex large-scale system. SD combined with econometrics, artificial intelligence, complex network, game theory and other disciplines is a research direction with the extensive and deepening applications of system dynamics.

References

1. Forrester, J.W.: *Industrial Dynamics: A Major Breakthrough for Decision Makers*. Harvard Business Review 36, 37–66 (1958)
2. Qun, L., Hong, M., Bin, X., Xiaoxue, W.: *Inverse System Simulation Study of Regional Population, Environment and Economy Coordinated Development*. J. Population Science of China 21, 193–199 (2005)
3. Bangzhao, W.: *Regional Industrial Economy System Dynamic Model and Policy Suggestions*. J. Statistics and Decision 10, 21–22 (2001)
4. Weixiong, Z., Guilin, X., Huilin, L.: *An Efficiency Evaluation of the Transportation-Territorial Economics Compound System*. J. Systems Engineering 5, 62–67 (2007)
5. Shi, Z., Wen, Z.: *Estimation of ARX Parametric Model in Regional Economic System*. J. Mathematics in Practice and Theory 9, 6–11 (2007)

6. Xiaocong, L., Manli, H.: The Application of Multilayer Distributed Intelligent Decision Support System to Regional Economy. *J. Microelectronics & Computer* 12, 172–174 (2006)
7. Xiaoming, C., Zhengbo, Y.: Study on Sustainable Development Early Warning System in Regional Economic Subsystem. *J. Ecological Economy* 11, 40–43 (2004)
8. Shenzhen Municipal Statistical Bureau. *Statistical Yearbook of Shenzhen*. Shenzhen Municipal Statistical Bureau (2006)

Industrial Connection Analysis and Case Study Based on Theory of Industrial Gradient

Zhi Yang, Wei Zeng, Hongtao Zhou, Ying Li, and Qi Fei

Institute of Systems Engineering, Huazhong University of Science and Technology,
Wuhan 43007, China
zengwei@mail.hust.edu.cn

Abstract. The difference and imbalance of regional economic development is an objective fact existed in the process of the socio-economic development. Giving an example of Wuhan city circle, the paper calculates and analyzes industrial gradient based on theory of industrial gradient. According to the characteristics of industries, the relevant industrial connection patterns are proposed, which provides the theoretical foundation for making development policy of regional industries.

Keywords: Industrial gradient, Industrial connection, Connection pattern.

1 Introduction

The difference and imbalance of regional economic development is an objective fact existed in the process of the socio-economic development. The difference between regional labor force and resources inevitably leads to the difference of regional economic development. The imbalance of regional economic development is especially obvious, such as the imbalance of regional economic development between the eastern coastal areas and the Midwest in China [1]. Even different cities in the same city circle have great differences in economic development, for example, Wuhan city has obvious advantages in economic development in Wuhan city circle, compared with other cities like Huanggang city. The difference in regional economic development in China has the trend to become wider gradually.

National and regional competitiveness is embodied through comprehensive competitiveness of national and regional industries. Industries are closely associated with regions, and the development of a region can't be separated from industries. After years of unbalanced development in China, there are obvious industrial and economic gradients in domestic regions. At the same time industry transfer among domestic regions is increasing and bringing about industrial upgrading, which means industries in developed regions transfer to underdeveloped regions. So, the study of industry transfer and industrial connection in Chinese regional economy is of great significance both in theory and practice.

Based on the theory of industrial gradient and a comparative study of industrial structure in Wuhan city circle, the paper studies connection patterns and suggestions for economic cooperation and coordinated development of Wuhan city circle.

2 Industrial Gradient Theory

Because of difference in production factors, industrial foundation and industrial division, there are gradients in regional economic development, industrial structure and technological level. The industrial gradient makes industry transfer possible [2].

Actually, gradient reflects the relatively high or low potential of each industry in economic space. Because of the close connection among most industries, economic cooperation among different regions refer to not only the industrial transfer caused by the gradient of the same industry in different regions, but also the changes of industrial structure caused by changes of different industries' industrial position, role, impact in the economic space.

Industrial gradient coefficient is defined to measure the gradient of regional industrial development [3]. The industrial gradient coefficient is mainly determined by two factors: one is innovation factor, which is expressed by comprehensive comparative labor productivity, and it depends on employee's skills, technological innovation level and their ability of converting to production in a region. Another is industrial concentration factor, which is expressed by the rate of production specialization, and it depends on factors such as the utilization degree of natural resources and the number of professional equipments and professional technical personnel in the industry.

2.1 Comparative Labor Productivity

Comparative labor productivity (CLP) is also called relative national income, denoting a comprehensive index of an industry's comparative advantages in a region. CLP represents the technological innovation factors of an industry, and embodies the competitive ability of an industry. Its formula is shown as follows:

$$\text{CLP} = \frac{\text{ratio of added value of an industry in a region to added value of the industry in its country}}{\text{ratio of employees of a certain industry in a region to employees of the industry in its country}} \quad (1)$$

Table 1. The comparative labor productivity of five cities in Wuhan city circle in 2006

	Wuhan	Huanggang	Huangshi	Xiaogang	Xiantao
agriculture, forestry, animal husbandry and fishery	1.679	0.775	0.792	0.947	0.791
industry	1.227	1.054	0.815	0.560	0.554
construction industry	1.601	0.615	0.300	0.016	15.567
post, warehousing and transportation industry	0.149	4.302	9.043	3.483	2.442
wholesale retail	1.022	0.617	1.298	0.752	1.799
lodging and catering trade	1.017	0.693	1.105	0.442	2.908
finance	1.610	0.051	0.220	0.031	0.174
real estate	1.139	0.319	1.561	0.715	1.084

2.2 Industrial Specialization Rate

Industrial specialization rate is a basic index of distinguishing the patterns of regional division, used to explain the regionalization level of a certain industry in regional division. Comparing industrial regionalization among different regions can show the basic pattern of territorial division, it is an indicator of analyzing regional industry layout and industrial advantages in modern economics. Industrial specialization rate is a proportion of two ratios; one is a ratio of output value of an industry to output value of all the industries in a region, while another is a ratio of output value of the industry to output value of all the industries in its country or province. The formula is shown as follows:

$$Q = \frac{Y_{a_i} / Y_a}{Y_i / Y} \tag{2}$$

Where:

Q is the output value of the specialization rate;

Y_{a_i} is the output value of industry i in region a ;

Y_a is the output value of all the industries in region a ;

Y_i is the output value of industry i in its province;

Y is the output value of all the industries in its province.

When $Q > 1$, it shows that the industry’s specialization degree is higher than the province level in the region. It means that production of this industry is concentrated in this region with the comparative advantage. The higher the value of Q, the higher the specialization degree and the larger the advantage will be. It also means that the industry’s output in this region not only meets the needs of this region, but also provides products or services for exterior regions. Comparing two different regions, e.g. region a and b , the index shows that the professional level of region a is higher than that of region b , and the industry’s development of region a is more preponderant than that of b .

Table 2. The industry specialization rate of five cities in Wuhan city circle in 2006

	Wuhan	Huanggang	Huangshi	Xiaogan	Xiantao
agriculture, forestry, animal husbandry and fishery	0.438	3.029	0.766	2.410	2.156
industry	1.031	0.683	1.286	0.866	0.896
construction industry	1.112	0.981	0.603	0.851	0.627
post, warehousing and transportation industry	1.083	0.645	1.403	0.666	0.359
wholesale retail	1.053	0.784	0.908	0.726	1.585
lodging and catering trade	1.160	0.662	0.822	0.500	0.959
finance	1.403	0.232	0.258	0.196	0.289
real estate	1.027	1.093	0.830	1.026	0.711

To identify the leading industry, only those industries whose specialization rate is over 1 can constitute the foundational industries of the region and play a leading role in the local economic development. When $Q < 1$, it means that the industry's specialization degree of the region is lower than that of country, and its scale is disadvantaged. The smaller the value of Q , the lower industry specialization will be, which means that the industry's output in this region can not meet the needs of the region, so it needs products or services provided by exterior region.

2.3 Analysis of Industrial Gradient Coefficient

The concept of industrial gradient comes from the concept of gradient in regional economics. From the perspective of regional economics, gradient is a representation of regional economic development gap on the map. The regional industrial gradient level is measured by the product of the rate of industry specialization and the comparative labor productivity, and it is called industrial gradient coefficient. The formula is shown as follows:

$$\text{Industrial gradient coefficient} = \text{Industrial specialization rate} * \text{Comparative labor productivity} \quad (3)$$

From the definition, the industrial gradient is codetermined by industrial concentration rate and labor productivity, which play a multiplier effect. That is to say, the higher the degree of specialization, the higher the labor productivity is. So their products are used to measure the size of industrial gradient. The method of the industrial gradient coefficient can make up the deviation caused by the inter-regional differences in labor productivity, which can't be expressed accurately by specialization rate. It is easy to get the comparative labor productivity data and calculate the coefficient.

Table 3. The industrial gradient coefficient of five cities in Wuhan city circle in 2006

	Wuhan	Huanggang	Huangshi	Xiaogan	Xiantao
agriculture, forestry, animal husbandry and fishery	0.735	2.348	0.606	2.282	1.705
industry	1.264	0.720	1.048	0.485	0.497
construction industry	1.780	0.604	0.181	0.013	9.760
post, warehousing and transportation industry	0.162	2.776	12.684	2.319	0.877
wholesale retail	1.076	0.484	1.178	0.546	2.852
lodging and catering trade	1.180	0.459	0.909	0.221	2.788
finance	2.258	0.012	0.057	0.006	0.050
real estate	1.170	0.349	1.296	0.733	0.770

From Table 3, there is certain difference in the development level of industries among these cities. Generally speaking, the proportion of the secondary and tertiary industry of Wuhan is bigger and relatively more developed, but at the same time, the primary industry and the warehousing and transportation industry of the other cities are relatively more developed.

From the viewpoint of difference in industry development, although the number of employees of Wuhan city's primary industry is relatively larger, the secondary industry is in a dominant position, and it is in the middle stage of industrialization at present. The other cities in Wuhan city cycle are different from Wuhan city. The output value of their three industries almost is same, and the employee is centralized in the primary industry, and the economic development level and the industrial structure are approximately in the initial stage of industrialization. On the whole, driven by production costs, labor costs and market, some of the traditional manufacturing industries in Wuhan need to transfer to other cities in Wuhan city circle, and at the same time, these cities also have basic industry conditions to accept these industries.

3 Industrial Connection Mode

From the viewpoint of industry level, industry transfer begins with labor-intensive industries such as textiles, and then capital-intensive industries such as iron and steel, petrochemical, metallurgical, and lower-level technology-intensive industries such as electronics and communication industries [4]. From the viewpoint of region, it is often transferred from developed regions to sub-developed areas, and then to undeveloped areas. The industry transfer is focused on the secondary industry. Some tertiary industries also expand outwards, such as transportation, trade services, financial insurance and so on. Because of the primary industry's own characteristics, there are huge obstacles to transfer.

3.1 Industrial Gradient Transfer Mode

Because of the imbalanced development in economic and technology, an economic and technological gradient is formed objectively. The differences in regional economic and technological gradient will give an impetus to economy, technology and productivity transfer on the space. According to gradient transfer rules of productivity, firstly, the regions of high industrial gradient introduce and master the advanced technology; and then transfer to regions of the secondary, third level gradient gradually [5]. With the economy developing, the speed rate of transfer accelerates, in order to narrow the gap in different regions and achieve a relatively balanced distribution of economy, and then achieve a balanced development of national economy.

For Wuhan city circle, the gradient transfer modes mainly apply to heavy industries with high technology, such as electronic communication, electric machinery manufacturing and transportation equipment manufacturing. These industries have a certain amount of demands in the cities of low industrial gradient like Huanggang city. In these cities these industries have the disadvantages of technology, capital, and production scale and so on. So these industries in Wuhan city have the trend to transfer to these cities of low industrial gradient gradually with the adjustment of industrial structure.

Industrial gradient transfer will bring about the development of related industries in low-gradient regions as well as optimize the industrial structure in high-gradient regions, and promote the industrial upgrading. The industrial cooperation in the industrial gradient transfer is effective in high technological industries and labor-intensive industries in the field of heavy industry, but less effective in the capital-intensive and knowledge-intensive industries.

3.2 Division and Cooperation Mode of Industrial Connection

The industry cooperation can happen in the same industry of different regions as well as in different industries [6].

Most of the economic cooperation in different regions with different economic development levels is the economic cooperation in different production stages, which is called vertical economic cooperation. The cooperation in Wuhan city circle is mainly based on vertical economic cooperation mode in the past years, which is a kind of vertical industrial division and has the characteristic of resources complementarity. In other words, other cities mainly product raw materials and provide Wuhan cities upstream products, while Wuhan is mainly engaged in the middle and downstream deep-processing industry. The vertical economic cooperation mode can be mainly divided into industrial chain expansion pattern, production in the place of sale pattern, as well as the commissioned processing pattern.

1) Industrial chain expansion pattern

The industrial chain expansion pattern is mainly applied to resource-dependent industries, such as coal, oil and natural gas industry. Take the coal industry as an example, its related industrial chain such as coal processing equipment; mine mining equipment, water supply and drainage tunnel ventilation equipment, instrumentation equipment and coal processing equipment can be developed, as well as the tertiary industry like the communication and transportation which is closely interrelated to the employee's lives. Through the expansion of industrial chain, the low-gradient region in Wuhan city circle can transfer from the mineral resources-oriented industrial structure to a comprehensive industrial structure, and promotes the regional economic development.

2) Production in the place of sale pattern

Production in the place of sale pattern is mainly applied to manufacturing industries in the field of light industry, such as food and beverage industry. These industries often have potential markets in low-gradient regions. However these industries almost aggregate in Wuhan city, where has the advantages of technology, capital and production scale. The advantage of having a large market in low-gradient regions is used to attract actively advanced enterprises in high-gradient regions to establish production bases and meet the local market's needs. This pattern is an effective means that achieve win-win in low and high industrial gradient regions.

3) Commissioned processing pattern

The commissioned processing pattern is mainly applied to labor-intensive industries such as textiles and paper industry. In Wuhan these industries have a long history of development and produce high quality products. However, because of high demand of labors, it makes such industries transfer to low industrial gradient regions like Huanggang city where labor force is abundant. Through the commissioned processing pattern the high and low industrial gradient regions achieve economic cooperation in city circle.

3.3 Horizontal Division Cooperation Mode

The cooperation in similar economic development level regions is usually horizontal economic cooperation [7]. In Wuhan city circle, the horizontal division cooperation

includes two patterns: one is brand transplant pattern, and another is association and merger pattern.

1) Brand transplant pattern

The brand transplant pattern is mainly applied to the cooperation of those industries which have been formed large brand effect. Competition is actually the brand competition in the market economy, and a famous brand in a world or nation is a valuable intangible asset. So in the different industrial gradient regions, brand transplant can create higher profits and promote common development of city circle's economy. In the example of Wuhan city cycle, brand transplant is that some well-known brands transfer in city cycle and realize localization production.

2) Association and merger pattern

The association and merger pattern is mainly applied to the cooperation of such industries as manufacturing industries with somewhat good industrial foundation. These industries have a long history of development in the low gradient region, but now they face some difficulties such as low management and technology level. Such industries with high gradient have strong development vigor due to its high capital and technical content. So the industry association and merger can bring about new development opportunities to those industries with low gradient.

4 Conclusions

Based on the theory of industrial gradient, the paper measures and analyzes industrial gradients through an example of Wuhan city circle. This paper points out that it is the existence of industrial gradient that makes the industry transfer possible and adjusts the industrial structure constantly. Several industrial connection patterns are discussed in the example of economic cooperation in Wuhan city circle, which provide a theoretical foundation for making regional industrial policy.

References

1. Taozhen, Y.: Industrial Migration and Chinese Regional Economic Gradual Development. Master's Thesis of Wuhan University (2006)
2. Hua, S.: Two Misunderstandings Concerning Theory and Practice of Industrial Transition. *Journal of Lanzhou University (Social Sciences)* 1, 137–140 (2001)
3. Rui, C., Biling, X.: The Strategic Conception of Regional Industrial Gradient based on the Improvement of the Industrial Gradient Coefficient. *Journal of Forum on Science and Technology in China* 8, 7–11 (2007)
4. Wenqi, T.: Research on the Industrial Pattern of Regional Economic Cooperation of the East and the West in China. Master's Thesis of East China Normal University (2006)
5. Zhongai, X.: A Study on Industry Gradient Transfer at the Region of Extensive-Triangle Zhu. *Journal of Hebei University of Science and Technology (Social Sciences)* 1, 16–22 (2006)
6. Rongqing, C.: How to Enhance Regional Industry Transfer and Structural Improvement. *Journal of Zhejiang Normal University (Social Sciences)* 4, 67–70 (2002)
7. Dongchen, Y.: An Exploration on Matching of Industries under Regional Economic Integration. Master's Thesis of Guangxi University (2007)

Extracting Schema from Semistructured Data with Weight Tag

Jiuzhong Li and Shuo Shi

Department of Computer Engineering, Guangdong Industry Technical College,
Guangzhou 510300, China
gzlijz@163.com

Abstract. This paper put forward the concept of OEM model with weight on its edges, develops a new approach to extracting schema from semistructured data with weight on its edges, and gives two theorems related to computing target set of label path and supporting degree of label path. Using width-first and top-down traversing strategy, the algorithm computes target set and supporting degree of every label in a label path, and decides whether the label is retained in schema model according to its magnitude of supporting degree and weight of the label. In the last, we test the validity and efficiency of the algorithm. The schema scale of the semistructured data obtained from the same OEM database in this paper is smaller than that in other paper.

Keywords: Semistructured Data, Schema extraction, OEM model with weight, Label path supporting degree, Target set of label path.

1 Introduction

With the large number of Web applications, lots of semistructured data is generated. Semistructured data is in the form of data between strict structured data and structure-free data (such as voice, image files). It has the following features [1-6]: (1) implicit schema information, (2) irregular structure, (3) no strict type constraint, but expresses through data. Since semistructured data has no explicit schema, it brings great difficulties to data storage, quick query and optimization, which results in a low efficiency in querying, browsing and integrating Web data.

The aim of schema study is to extract the internal structures from semistructured data. There are two forms of schema descriptions, one is based on logic description, and the other is based on graphic description^[1-2]. Some methods based on the graphic description were discussed in [3-6]. But the algorithms are complex, and the resulting schema is so large that their size even exceeds the original semistructured data.

In the algorithms proposed in [3, 5], label path with lower supporting degree would be pruned. However, some label paths contain indispensable information for the schema. To avoid pre-pruning such label paths we raise a new method in this paper, which extracts schema from semistructured data with weight. Comparing with those in [3-6], this new method obtains smaller-scale schemas with a faster algorithm, overcomes their disadvantages.

paths according to the minimum support degree, but also selecte the label paths according to the weight of edge contained in label path.

2.2 Terms and Theorems on OEM with Weight and Schema Extracting

Definition 1 ^[4, 6]. A label path of an OEM database G is a sequence of dot-separated labels, such as $l_1.l_2 \dots.l_n$. We can traverse a path of n edges e_1, \dots, e_n , starting from the root node where edge e_i has label l_i , the corresponding label path is $lp= l_1. l_2. \dots. l_n$

Definition 2 ^[4, 6]. A data path dp of an OEM database G is a alternating sequence of object identifiers and labels, written as $dp= O_0, l_1, O_1, l_2, \dots, l_n, O_n$. We call n the length of the dp. A $dp=O_0, l_1, O_1, l_2, \dots, l_n, O_n$ is called a instance of the label path $lp= l_1. l_2. \dots. l_n$. For a lable path lp there may exist more than one instances.

Definition 3 ^[6]. In an OEM database G, a target set t of a label path $lp= l_1. l_2. \dots. l_n$ is defined as the set of the last object identifiers of all data path instances, that is, $t=\{O | O_0,l_1 ,O_1,l_2, \dots, l_n,O$, where $O_0,l_1 ,O_1,l_2, \dots, l_n,O$ is a instance of the label path lp}. We also write $t=T(lp)$.

Definition 4 ^[5]. Let $lp= l_1. l_2. \dots. l_n$. The number of data path instances of lp is called the supporting degree of lp, denoted by Sup(lp).

We set a minimum supporting degree threshold, denoted by minsup, in order to remove low-frequency label path. The label path is frequent if label path supporting degree is greater than or equal to minsup, otherwise, the non-frequent.

To showe some of label paths an important position in model, we introduce to the concept of the directed edge with weight.

Definition 5. In an OEM database G, we assign a number w as the weight to each directed edge. And hence an edge is now denoted by $\langle o_i, l_n[w], o_j \rangle$. For instance, a directed edge $\langle O, name[2], 30 \rangle$ has the weight 2. The default value for the weight of a directed edge is 1, and we skip the number when the directed edge has weight 1.

Now we can prove the following two theorems, which are related to target sets and supporting degree of label paths.

Theorem 1. *In an OEM database G, let O_1 and O_2 be two objects in the target set of a label path lp. And suppose that there are two directed edge $\langle O_1, l, O_i \rangle, \langle O_2, l, O_j \rangle$, then O_i, O_j belong to the target set of the same label path.*

Proof. Let $lp= l_1. l_2. \dots.l_k$ and the corresponding data paths for O_1 and O_2 are $dp_1=x_0, l_1, x_1, l_2, x_2, \dots, l_k, O_1$ and $dp_2=y_0, l_1, y_1, l_2, y_2, \dots, l_k, O_2$ respectively. Let $dp_1= x_0, l_1, x_1, l_2, x_2, \dots, l_k, O_1, l, O_i$ and $dp_2= y_0, l_1, y_1, l_2, y_2, \dots, l_k, O_2, l, O_j$. Then dp_1 and dp_2 are two data path instances of label path $l_1. l_2. \dots.l_k.l$. Therefore, O_i, O_j is two objects in the target set of the label path $l_1. l_2. \dots.l_k.l$. □

Theorem 2. *In an OEM database G, let $lp_n= l_1. l_2. \dots.l_n$ be a label path and $T(l_1. l_2. \dots. l_{n-1})=\{O_{i1}, O_{i2}, \dots, O_{ik}\}$, then the supporting degree of the label path lp_n is equal to the number of directed edges that has label l_n and starts from the objects in $\{O_{i1}, O_{i2}, \dots, O_{ik}\}$, that is, the number of elements in target set of l_n (repeated elements counted multiplication).*

Proof. According to the definition of label path supporting degree, the supporting degree of the label path $lp_n = l_1, l_2, \dots, l_n$ is equal to the number of data path instance. Let $sup(lp_n) = s$, and let the s data path instance of lp_n be $dp_i = x_{i0}, l_1, x_{i1}, l_2, x_{i2}, \dots, l_{n-1}, x_{in-1}, l_n, x_n$, where $i = 1, 2, \dots, s$. Then, there are s edges with label l_n started from the set $\{x_{1n-1}, x_{2n-1}, x_{sn-1}\}$ (including repeated elements). However, $\{x_{1n-1}, x_{2n-1}, x_{sn-1}\}$ is a subset of $T(l_1, l_2, \dots, l_{n-1})$, hence the number of directed edges of label l_n started from $T(l_1, l_2, \dots, l_{n-1})$ is at least s . But $sup(lp_n) = s$. So, the number of directed edges of label l_n started from $\{O_{i1}, O_{i2}, \dots, O_{ik}\}$ is exactly s , then, there are s elements in target set of label l_n (repeated elements count multiplication). \square

2.3 Storing Semistructured Data in OEM Database with Weight on Its Edges

Data storage structure is crucial for designing good algorithms. In order to rapidly compute the target set and supporting degree of a label path, we store the OEM database $G(r, V, E)$ with adjacency list method of directed graph deformation, the oids that start from the same oid and can be reached by the same label are stored together for rapidly computing. Stored node structure can be understood by the following example and we give no more explain. The head node and structure of a list is shown in Figure 2. The storage structure of OEM instance database with weight in Figure 1 is shown Figure 3.

The head node of list is stored into a linear table, so that any node and their children nodes can be randomly accessed. So we can rapidly compute the target set of any label and the label sets which started from the target set.

Oid	link
-----	------

Label	End_oids	link
-------	----------	------

The type of head node in chained list

The type of the node in chained

Fig. 2.

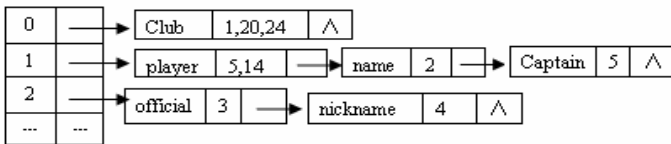


Fig. 3. OEM Examples of database storage structure with weight

3 Schema Extraction

Schema extraction is to find out the inner schema picture from semistructured data, that is, the set of maximal label paths that appear with high frequency or have higher weights. To describe OEM data schema with weight, we establish a schema table. The node structure in the table is represented as (father,label,weight,T_objs,child), where father and child are both schema node Oid, and there is a directed edge with label connection from father to child, T_objs is the target set of the label, mapped to the child node in schema table, and weight means the weight of the label, with default value 1.

In order to solve the problem which the target set contain the atom type, we use the sign ‘ \perp ’ to indicate a special target set of some label path which may contain atomic object of Oid.

3.1 The Ideas of Schema Extraction

The ideas of schema extraction are as below. Firstly, we map the root node of OEM database G with weight to the root node of the schema. Then, starting from the root node r of G , using width-first and top-down traversing strategy, we find out the label set that the label started from the root r , the target set of every label corresponding to the directed edges and the supporting degree of the label paths that contains the label. When the supporting degree of the label path is less than the minimal supporting degree and the weight of the label is less than minweight , this label will be pre-pruned from the label paths; when the weight value of the label is greater than or equal to minweight , or the label support degree is greater than or equal to minsup , we generate a composite node related to this label, node_y ($\text{father}, \text{label}, \text{T_objs}, \text{sup}$), and store node_y into a temporary array temp . Then, we sort the elements in array temp by value of its supporting degree in descending order, and examine T_objs field of the elements to see whether which is subset of some node’s T_objs in the schema table. If the answer is ‘yes’, we add directed edge with label from father node to the schema node which T_objs maps to the node; otherwise we generate the new schema node. Then, we put the corresponding element in temp array into queue (Q). While the queue is not empty, the head element of temp sends out Q , repeats the above process to compute the label set, label supporting degree starting from the target set T_objs , and then, decides whether some label should be pruned by the same standard. The process is repeated until the queue is empty.

3.2 The Algorithm for Schema Extracting with Weight Tag

Algorithm 1. An Algorithm of Schema Extracting with Weight Tag

Input: The root node r and their label root_label of graph G , minimal support degree minsup , minimal weight minweight

Output: Schema table Schema_table

```

Schema_extracting( $r, \text{minsup}, \text{root\_label}, \text{minweight}$ )
  {Initqueue( $Q$ ); // Initialization and the queue  $Q$  is set to be
empty
   $l_n = \text{root\_label}$ ; Count = 0;
   $\text{T\_objs} = \{ r \}$ ; // Target set of  $\text{root\_label}$  sets as  $\{r\}$ 
   $\text{Schema\_table.Insert}(\text{Generate\_point\_node}(\text{null}, \text{root\_label}, \text{T\_objs}, \text{count}))$ 
  // mapping OEM root node into schema root node, then inserting
into schema table
  Father = count;
   $\text{Node}_x = \text{Newobject}(\text{father}, \text{T\_objs})$  // generating a new node
  AddQ( $Q, \text{Node}_x$ ) // compound node enters queue  $Q$ 
  Do while not empty( $Q$ )
  {  $\text{Node}_x = \text{DEL}(Q)$ ; // The head element of queue sends out queue

```

```

s_objs = Node_x.T_objs; Father = Node_x.father;
ls = labels_starting(s_objs); //The function returns to the
all label set which starting from s_objs
if(ls=null) continue;
Temp = { }; // Initializing the temporary array
for each  $l_n \in ls$ 
{ sup = 0; T_objs = target_objects( $l_n$  , s_objs, &sup);
//Return the target set and supporting degree which
starting from s_objs
 $l_n\_weight = 1$ ; //The default weight of label is 1
 $l_n\_weight = readweight(l_n)$ ; //Fetching the weight of label  $l_n$ 
If (sup >= minsup or  $l_n\_weight \geq minweight$ )
{node_y=newobject(father, $l_n$ ,T_objs ,sup)
//generating a temporary node which has 4 fields
insert into temp array; //insert into temp array in turn}}
sort elemnts of temp by value of temp[i].sup in descending
order
for each temp[i] // i from 1 to maximum
{d=schema_table.search(temp[i].T_objs,count)
//Searching for Schema_table, if temp [i].T_objs is the
subset of T_objs of some node in schema table, then returning
the schema node Oid which Corresponding this set, otherwise
returning null.//
if(d!=null and Schema_tabel[d].Label<> $l_n$ )
Schema_tabel.Insert(Generate_point_node(temp[i].Father,temp[i]
.Label, temp[i].T_objs , d) );
//insert a node into the schema ,adding a directed edge with
the label temp[i].label pointing to the schema node d//
Else
{ d = Generate_point_node (temp[i].Father, temp[i].Label ,
temp[i].T_objs, ++count); //target set temp[i].T_objs of
temp[i].Label is mapped to a new schema node d
Schema_tabel.Insert(d); //node d is inserted into Schema_tabel
Node_x = newobject(temp [i].Father,temp[i].T_objs)
AddQ(Node_x); //Enter queue } } }

```

Algorithm 2. Computing the label set which starts from the nodes in the set of target

```

Labels_starting(s_objs)
{ ls={ } //Initializing the label set
For each obj  $\in s\_objs$ 
Add each label name stored in linknode in the adjacency
list of adjlist[obj] into ls;
Return ls; }

```

Algorithm 3. Computing target set of label l_n and which starts from object of objs and support degree of label path lp which ends with l_n

```

Target_objects( $l_n$  , objs, *count_ $l_n$ )
{ // computing the target set T of label  $l_n$ , where *count_ $l_n$  is
the supporting degree of  $l_n$ 
*count_ $l_n$ =0; T = { };
for each obj  $\in$  objs and adjlink_list[obj]

```



```

{for each linknode of adjlink_list[obj]
{ IF ((*p).Label = l_n)
{//The p points to a node of adjacency list of adjlink_list[obj]
  for each oid ∈ (*p).End_oids
  { (*count_l_n)++; T =T ∪ { oid };
  If ( the oid node is atomic type )
    T =T ∪ { ' ⊥ ' };
    } } } }
Return T; }

```

3.3 Instance Analysis

Applying above algorithms to Figure 1 of OEM with weight sample database, we can obtain the schema table shown in Table 1. The schema graph for sup =1 , weight =1 is shown in Figure 4. However , to obtain the schema table for sup>=minsup or weight >=minweight, we only need to delete the row corresponding to sup<minsup and weight <minweight as well as the branch row for this row target set.

Table 1. Schema table for Minimum supporting degree 1

S_objs	father	Label	Weight	T_objs	sup	Child
	Null	premiership	maxsize	{ 0 }	maxsize	0
{ 0 }	0	Club	1	{1,20,24}	3	1
	0	Name	2	{30}	1	2
{1,20,24}	1	player	1	{5,14,22,28}	4	3
	1	Name	1	{2,21,25, ⊥ }	3	4
	1	captain	1	{5}	1	3
	1	stadium	1	{26}	1	5
{30}	2	first	2	{31, ⊥ }	1	6
	2	last	2	{32, ⊥ }	1	7
{5,14,22,28}	3	Name	1	{6,15,23,29, ⊥ }	4	8
	3	nationality	1	{12,19, ⊥ }	3	9
	3	number	1	{13,18, ⊥ }	2	10
	3	formerclub	1	{1,24}	2	1
.....

As Table 1 shows, the target set of label player is {5, 14, 22, 28} and the target set of caption is {5}, which is the subset of label player’s. So, in the algorithm we just add a directed edge with label captain pointing to schema node ③, which is consistent with Figure 1. Obviously, captain branch is redundant, because the structure of captain branch is the same with player’s. So our algorithm excel those in [4, 5], whose outcoming schema still have such a branch. Analogously, the formerclub target set is subset of club target set, so we add the directed edge with label formerclub pointing to schema node ①. So, we have solved schema extracting from semistructured data with loop structure and with weight on its directed edge.

Remark. A. As S_objs column in schema table is the same with T_objs, it can be removed in schema table. The root schema node is a special node; it needs to be generated in all case. Therefore, it is given the maximum value of Weight and sup. Corresponding to Figure 1, the schema graph for sup=3, weight =2 is shown in Figure 5.

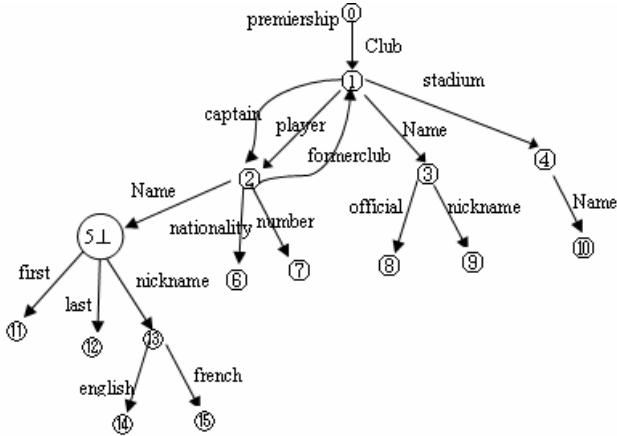


Fig. 4. Schema for supporting degree 1 ,minimum weight 1

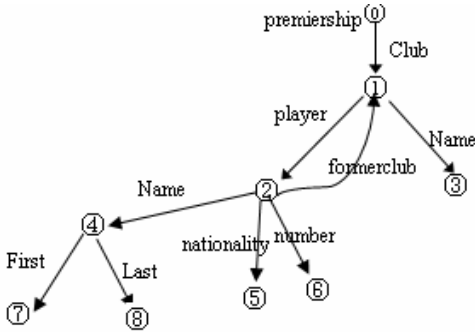


Fig. 5. Schema for supporting degree 3 ,minimum weight 2

3.4 Algorithm Validity

We use the data example of [5] to test the validity and efficiency of our algorithm. The result is shown in Table 2.

Table 2. Algorithm running time

running time		Algorithm in paper	Algorithm in [5]
Minsup	minweight		
Minsup=1	minweight=1	2055	4170
Minsup=500	minweight=1	10	17

The experimental result in Table 1 shows that the algorithm efficiency is significantly improved in our paper, comparing with [5]. And the algorithm can be used in wider range than those in [4-5]. The algorithms in this paper can apply to extracting schema from semistructured data with loop structure and with weight tag on its edges.

4 Conclusion

This paper proposes a new algorithm for extracting schema from semistructured data with weight and improves and extends the algorithms of schema extraction in [4-6]. The schema scale of the semistructured data for same database obtained in this paper is smaller than that in paper[4-6], and the algorithm can be applied in a wide range. There are still much to discuss on the study of semistructured data, such as semistructured data model, semistructured and materialized view and other dynamic view of maintenance.

References

1. Wang, J., Meng, X.F.: Schema of Semistructured Data. A Survey. *Computer Science* 2, 6–10 (2001)
2. Meng, X.F.: An Overview of WEB Data Management. *Journal of Computer Research and Development* 4, 385–395 (2001)
3. Lu, C., Wei, C.Y., Zhang, H.T.: Schema Discovery of Semi-Structured Data Based on OEM Model. *Computer Engineering and Applications* 34, 162–165 (2006)
4. Meng, D.L., Ye, F.Y., Li, X.H.: Extracting Schema from Semistructured Data. *Computer Engineering and Applications* 27, 162–165 (2006)
5. Liu, F., Hu, H.P., Lu, S.F.: Schema Discovery for Semistructured Hierarchical Data. *Mini-micro Systems* 1, 84–88 (2004)
6. Goldman, R., Idom, W., DataGuide, J.: Enabling Query Formulation and Optimization in Semistructured Databases. In: *Proc of the international Conf of the Very Large Data Bases(VLDB)*, Athens, Greece (1997)
7. McHugh, J., Widom, J., Loo: A Database Management System for Semistructured data. *SIGMOD Record* 3, 54–66 (1997)
8. Wu, G.Q., Chen, E.H.: An Approach of Storing Semi-structured Data with XML. *Computer Engineering* 10, 57–59 (2005)

Designing Domain Work Breakdown Structure (DWBS) Using Neural Networks

Yongjun Bai, Yong Zhao, Yang Chen, and Lu Chen

Institute of Systems Engineering, Huazhong University of Science and Technology, Wuhan,
430074 Hubei, China
yongjunbai@yahoo.com.cn

Abstract. Work breakdown structures are the basis of project management. But there are few researches on the methods or tools to design work breakdown structures effectively. In this paper, a framework which employs neural networks to plan the work breakdown structures has been introduced. The main concepts, including domain tree structure(DTS), domain work breakdown structure(DWBS) and relational work breakdown structure(RWBS), have used to form the outputs of the model. The nature of projects, which have been represented by a limited set of attributes, are considered as the main inputs of the model. Since the work breakdown structure is a hierarchical structure, DWBS has been broken into levels to reduce the complexity of reasoning and calculation. In addition, to make sure the result WBS is optimized and can be mapped into other WBSs effectively, Axiomatic Design Theory was used to verify the RWBSs at each level.

Keywords: Domain work breakdown structure, Neural network, Domain tree structure.

1 Introduction

Nowadays, projects become so large and complex that it is very difficult to manage the projects effectively. Especially, project planning is facing a huge challenge. Project planning is one of the most important steps in the process of project management, which impacts many of the steps to follow. The main goal of project planning is to develop the project work plan, considering various aspects. The project work plan could be used to predict the situation of project during its lifecycle, and enables the control of its progress trend.

Many researches were mainly about developing mathematical models for time scheduling and resource assignment, which should be done after the planning of work breakdown structure(WBS). Despite of WBS's importance and profound effects, there are a few researches about the methodologies or tools to develop the appropriate work breakdown structure for a given project. Most of the related works are only usable in special conditions and for very special kinds of projects. Then the planner should apply trial and error and unknown methods, to find the work breakdown structure and activities of a given project, which could increase the total risk.

This paper is going to introduce a process which shows how the work breakdown structure (WBS) and the activities of projects could be developed, with the aid of neural networks.

2 Related Works

There has been a lot of research on developing network-based planning methods and project management techniques, assuming that the activities and network structure are readily available for the project manager. Results of Tavarez's reviews shows that most of the related researches were mainly based on tree subject categories including project modeling, project evaluation and project scheduling and monitoring[1]. He noted that most of the researches were about developing mathematical models in resource assignment or project scheduling for different kind of projects. Meanwhile only some simple activity and network-based modeling techniques were stated, and there were not any procedures or even prescriptions about how to recognize the activities of projects.

Recently some emphasis has been placed on implementing neural network technology, in the development of automated process planning systems. Knapp and Wang formulated an approach for the automated acquisition of process selection and within-feature process sequencing knowledge using neural networks[2]. It is believed that human project planners will have more problems in generating the work breakdown structure, activities and network plan of projects, without referring to previous cases. The main reason for this is the huge amount and very complex of required knowledge. Therefore it is expected that project planners will try to search within previous project plans, and modify the most similar, to generate the work breakdown structure and activities of the current project.

From the view of decision making, the planning of project work breakdown structure and activities is a decision problem which should be solved by the project planner. Therefore some kind of decision support system could be applied to manage the knowledge of previous decisions and use them to support the planning of future projects. As the knowledge seems to be very complex, the system should be some kind of hybrid or intelligent decision support systems, which has been discussed and developed for some other decision problems[3,4].

Hashemi, et al. have presented a process to design WBS with the aid of neural networks[5]. They classified the WBSs of a project into three categories: Project Control Work Breakdown Structure(PCWBS), which shows the overall components of project main deliverable or subject; Functional Work Breakdown Structure(FWBS), which shows the scope of functions and operations; Relational Work Breakdown Structure(RWBS), which shows the relationship between PCWBS and FWBS. They also presented a modular neural network to generate PCWBS, FWBS and RWBS for a project. But their research has some main defects. A project usually contains several domains and the WBS used within a specific domain may be different from other domains'. Because of the projects' scale and complexity, the number of components of WBSs is too huge to be calculated effectively as a whole. Using neural network is only to generate a WBS for a specific project. Whether the resulted WBS is the best or not has not been verified by the authors.

3 Project Structures

3.1 Domain Tree Structure (DTS)

A domain is a collection of activities, which are launched by a specific group of experts in a particular environment for the completion of specific goals and tasks. Domain’s definition. An equipment acquisition project usually contains domains of function requirement(FR), design parameters(DP), process variables(PV) and maintenance processes(MP). Activities within a specific domain are relatively independent of activities of other domains. But the domain itself usually has relationship with other domains. We can illustrate these relationships by a tree. Within the domain tree structure, each parent domain has one or more children domain and each child domain can only has one parent. Parent domains defines the targets for their children domains and Children domains present the solutions for their parent domains. Fig. 1 shows the sample domain tree structure for an equipment acquisition project.

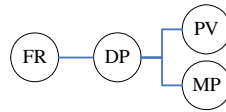


Fig. 1. An equipment acquisition project’s domain tree structure

In Fig. 1, FR is the root of the tree structure and has no parent domain. In other words, a project usually starts with requirement analysis. FR has one child of DP. So FR defines the targets that DP must reach. DP has two children, PV and MP. So PV and MP provide solutions for DP from different aspects.

Table 1. A DTS’s Adjacency Matrix

D	FR	DP	PV	MP	CC
FR	0	1	0	0	1
DP	0	0	1	1	2
PV	0	0	0	0	0
MP	0	0	0	0	0
PC	0	1	1	1	---

We have implemented the equivalent adjacency matrices of DTS. In general, let $DTS = (D; E)$ be an undirected graph with vertices $D = \{D_d, 1 \leq d \leq m\}$ and edge set E . D is the set of domains and d is the ID of a specific domain; m is the maximum number of domains within a project. The adjacency matrix $A = [a_{ij}]$ of DTS is the $m \times m$ symmetric matrix in which $a_{ij} = 1$ if and only if $i \neq j$ and D_i is adjacent to D_j , which means there is an edge between D_i and D_j .

$$CC_i = \sum_{j=1}^m a_{ij} \text{ means the children number of domain } i \text{ and } PC_i = \sum_{j=1}^m a_{ji} \text{ means}$$

the parents number of domain i . Table 1 shows a DTS’s adjacency matrix that corresponds to Fig. 1.

3.2 Domain Work Breakdown Structure (DWBS)

Even though working with the same system towards the same goal, experts from the different domains use their own specific tools, providing their own specific views of the system to be developed. So different work breakdown structures (WBS) are used in different domains. In fig. 1, the first step is to determine the Customer Needs or attributes that the system must satisfy and to create the FR structure to satisfy the customer needs. The next step is to map the FR structure into the DP domain – conceiving a design embodiment and identifying the DP structure. In this case, the WBS of a specified domain is not independent of the WBSs of other domains. They must be mapped each other.

The Domain Work Breakdown Structure(DWBS) is the hierarchical structure which shows the overall components of project main deliverable or subject within a specific domain. Fig. 2 and 3 illustrate the DWBSs of FR and DP domain of the project of “Acquisition of Missile” as an example.

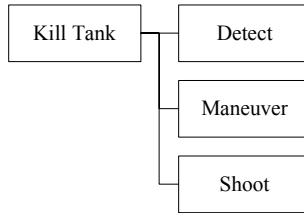


Fig. 2. The FR DWBS of project of Acquisition of Missile

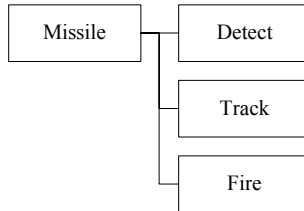


Fig. 3. The DP DWBS of project of Acquisition of Missile

Since DWBS is a hierarchical structure, it consists of components from different levels. DWBS is usually designed through a top-down procedure. In other words, the first level of DWBS is designed first and then the second level and then the third level, etc. And the components at a level must be mapped to the components at the same level which belong to other domains. So we breakdown the DWBSs into vectors

according to the components' level, $DL_{d,l} = \{v_1, v_2, \dots, v_{n_l}\}$. $DL_{d,l}$ means DWBS's level vector of domain d at level l ; l means level ID; v_i means component i ; n_l means the maximum number of components at level l . The DP DWBS in Fig. 3 can be broken into two vectors, $DL_{2,1} = \{Missile\}$ and $DL_{2,2} = \{Detect, Track, Fire\}$.

Since the matrix form is more suitable to be used as the inputs and outputs of neural networks, we have implemented the equivalent adjacency matrices of DWBS. Because of the huge number of DWBS's components, we break the adjacency matrices into leveled matrices. In general, let $DL_{d,l-1}$ be the rows and $DL_{d,l}$ be the columns. The adjacency matrix $A_{d,l} = [a_{ij}]$ is the $n_{l-1} \times n_l$ matrix in which $a_{ij} = 1$ if and only if v_i is adjacent to v_j , which means there is an edge between v_i and v_j . The design DWBS in Fig. 3 can be illustrate by the matrices in table 2.

Table 2. Design DWBS's adjacency matrices

	Detect	Track	Fire
Missile	1	1	1

3.3 Relational Work Breakdown Structure(RWBS)

The Relational Work Breakdown Structure(RWBS) represents the relationships between the components of different DWBSs. In other words, RWBS shows what components of a DWBS should be performed to obtain each component of the adjacent DWBS. The RWBS should be presented by a matrix, where the components of a DWBS are considered as the columns, and the components of the parent DWBS as the rows. If any part of the DWBS was in the process of achieving a certain part of the parent DWBS, the cell which is located in the intersection of the related row and column should be assigned to (1), otherwise to (0). Each RWBS is for a specific level. So RWBS was defined as $RWBS_{i,j,l}$. The i and j are Domain ID; l means level ID. Table 1 corresponds to a part of the RWBS between requirement DWBS and design DWBS at level 1, which has been considered for the project of "Missile Acquisition". The matrix was showed in Table 3.

Table 3. The RWBS between requirement DWBS and design DWBS at level 2

	Detect	Track	Fire
Detect	1	0	0
Maneuver	0	1	0
Shoot	0	0	1

3.4 Project Attributes

Project attributes are a limited set of variables which describes the characteristics of projects. We have defined two major sets of project attributes. The first is the general set of attributes(PA), which is applicable in describing the whole project instead of domains within the project, such as the project main domains, the project main subject, total amount of time, total amount of budget, maximum amount of expected risk, etc.. The second is the specific set of attributes(PA_d , d means domain ID), which is applicable in a specific domain, such as domain’s main subject, domain’s time, domain’s budget, etc..

4 The Methodology

The overall process, which is proposed to model the relationships between project attributes and structures with the aid of neural networks, is presented in Fig. 4.

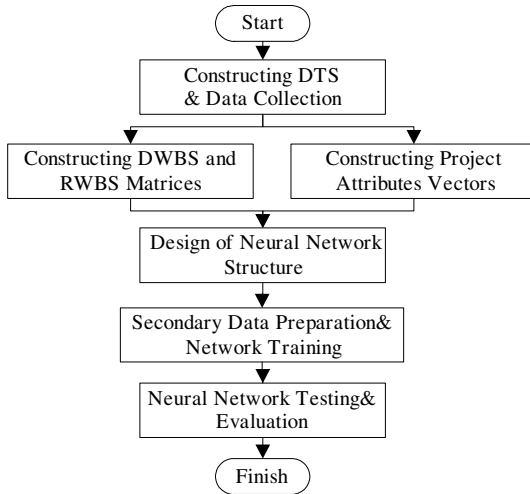


Fig. 4. The overall process

① Constructing DTS and Data Collection

The main objective of this step is to collect sufficient amount of data from the previous projects and to identify the domains of project. The projects’ domain tree structures must be created during this step.

② Constructing DWBS and RWBS Matrices

After identifying the domains of projects, we have constructed the DWBS for each domain at each level and RWBS between domains with respect to the collected data.

③ Constructing Project Attributes Vectors

In this step, we have constructed the global project attributes and domain project attributes for each domain.

④ Designing the structure of neural network model

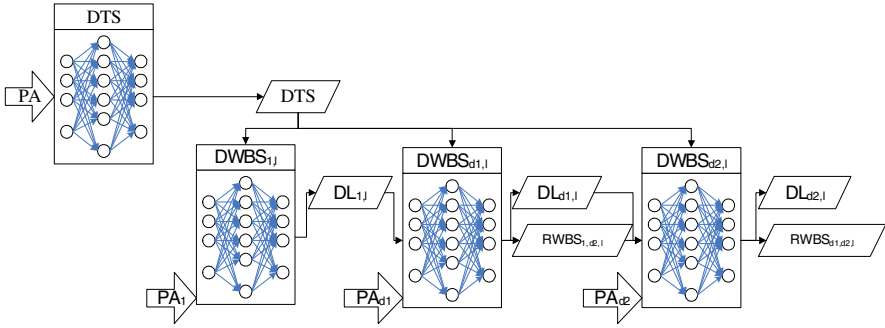


Fig. 5. The overall structure of proposed neural network

In this step, the structure of the proposed neural network which should be implemented for learning the relationships between projects attributes and their structures should be designed. In this regard, we have used a hierarchical neural network.

The first network is DTS with the input of the global project attributes. Its output is project’s DTS matrix. The DTS matrix is also the input of other networks to describe the domains’ relationships.

The set of networks for *DWBS* is for a specific level. The level starts with 1 and end with the maximum level number. While a specific level’s training finished, a higher level’s training begins. Since the root domain of DTS has no parent, so a special network is designed, which is named $DWBS_{1,l}$. $DWBS_{1,l}$ has inputs of DTS and domain project attributes($PA_{1,l}$) and generates $DL_{1,l}$. The successor DWBS networks have inputs of DTS, domain project attributes, DL and RWBS.

⑤ Secondary preparation of data and training the neural network

After completing the design of neural network model, we have to process the collected data, to prepare the training and testing patterns with respect to the final structure of the modules.

⑥ Testing and evaluating the neural network

In this step, the neural network modules should be tested and evaluated by sample projects.

In addition, each network’s output must be evaluated to assure the design result is the best for the project. Using Axiomatic Design Theory[6,7], the relationships between two adjacent domains’ WBSs at a specific level was illustrated in formula (3).

$$DL_{d_1,l} = RWBS_{d_1,d_2,l} \bullet DL_{d_2,l} \tag{1}$$

Whether the resulted DWBS is optimized or not depends on the corresponding RWBS’s form. According to the axiomatic design theory, optimized RWBS usually has two typical forms: diagonal matrix and triangular matrix. If RWBS is diagonal matrix, $DL_{d_2,l}$ is an uncoupled solution for $DL_{d_1,l}$. And if RWBS is triangular matrix, $DL_{d_2,l}$ is a decoupled solution for $DL_{d_1,l}$. So during the evaluation of the neural networks, the resulted RWBSs must be in the form of diagonal matrix or triangular matrix.

5 Conclusions

Above all, we have decomposed a project into domains and created domain tree structure. In the DTS, the relationships among domains have been classified as parents and children, or targets and solutions. Different domains use different work breakdown structures, which is defined as DWBS. And The WBS of children domain should be mapped into the one of parent domain. The mapping process is at each level. So we have decomposed DWBS into levels(DLs) and created the corresponding relationship matrices(RWBSs) at a specific level. We presented a neural network model to help designing the project's WBS domain by domain and level by level. At last, axiomatic design theory was used to evaluate the effectiveness of the resulted WBSs.

References

1. Tavarez, L.V.: A Review of the Contribution of Operational Research to Project Management. *European Journal of Operation Research* 136, 1–18 (2002)
2. Knapp, G.M., Wang, H.S.: Acquiring, Storing, and Utilizing Process Planning, Knowledge Using Neural Networks. *Journal of Intelligent Manufacturing* 3, 333–344 (1992)
3. Turban, E.: *Decision Support Systems and Intelligent Systems*, 6th edn. Prentice Hall, London (2000)
4. Zeleznikow, J., Nolan, J.R.: Using Soft Computing to Build Real World Intelligent Decision Support Systems In Uncertain Domains. *Decision Support Systems* 31, 263–285 (2001)
5. Hashemi, G.S., Emamizadeh, B.: Designing Work Breakdown Structures Using Modular Neural Networks. *Decision Support Systems* 44, 202–222 (2007)
6. Suh, N.P.: Axiomatic Design Theory for Systems. *Research in Engineering Design* 10, 189–209 (1998)
7. Suh, N.P.: *Axiomatic Design: Advances and Applications*. Oxford University Press, USA (2001)

Practical Hardware Implementation of Self-configuring Neural Networks

Josep L. Rosselló, Vincent Canals, Antoni Morro, and Ivan de Paúl

Electronic Systems Group, Physics Department, Universitat de les Illes Balears (UIB),
07122 Palma de Mallorca, Spain
j.rossello@uib.es

Abstract. This work provides practical guidelines for an efficient hardware implementation of Neural Networks. Networks are configured using a practical self-learning architecture that iterates a basic Genetic Algorithm. The learning methodology is based on the generation of random vectors that can be extracted from chaotic signals. The proposed solution is applied to estimate the processing efficiency of Spiking Neural Networks.

Keywords: Neural Networks, Spiking Neural Networks, Hardware implementation of Genetic Algorithms.

1 Introduction

The development of efficient solutions for the hardware implementation of neural systems is currently one of the major challenges for science and technology. Due to their parallel-processing nature, among other applications, neural systems can be used for real-time pattern recognition tasks and to provide quick solutions for complex problems that are intractable using traditional digital processors [1,2]. The distributed information processing of Neural Networks also enhances fault tolerance and noise immunity with respect to traditional sequential processing machines. Although all these processing advantages, one of the main problems of dealing with neural systems is the achievement of most advantageous network configurations since network complexity increases exponentially with the total number of connections. Therefore, the development of learning strategies to quickly obtain optimum solutions when dealing with huge network configuration spaces is of high interest for the research community.

Recently, a lot of research has been focused on the development of Spiking Neural Networks (SNN) [3] as they are closely related to real biological systems. In SNN information is codified in the form of voltage pulses called Action Potentials (APs). At each neuron cell the AP inputs are weighted and integrated in a single variable defined as the Post-Synaptic-Potential (PSP). The PSP is time dependent and decays when no APs are received. When input spikes excite the PSP of a neuron sufficiently so that it is over a certain threshold, an Action Potential is emitted by the neuron and transmitted to the rest of the network.

In this work we present a practical implementation of SNN. We also develop a simple architecture that can be used for SNN self-configuration. The proposed

self-learning solution has been applied for temporal pattern recognition. The rest of this paper is organized as follows: in section 2 we show the proposed SNN self-learning architecture, while in section 3 we apply the proposed system to temporal pattern recognition analysis. Finally in section 4 we present the conclusions.

2 Digital SNN Architecture

2.1 Digital Spiking Neuron Model

As mentioned in the previous section, in SNN the information is codified in the form of pulses. We used a simplified digital implementation of the real behavior of biological neurons. In the proposed system, the PSP decay after each input spike is selected to be linear instead of the real non-linear variation while the refractory period present after each spike emission has been neglected. The main objective of this work is not to provide an exact copy of the real behavior of biological systems but to develop a useful Neural Network configuration technique. In Fig. 1 it is shown an example of the dynamic behavior of the digital model implemented.

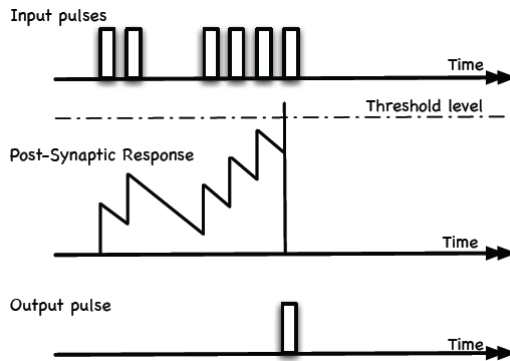


Fig. 1. Dynamic behavior of the digital implementation of spiking neurons. Action potentials are represented as digital pulses while post- synaptic potential variation is assumed to be linear with time. The typical refractory period after each output spike has been neglected in this model.

In the digital version implemented the PSP is codified as a digital number. At each spike integration the PSP is increased a fixed value that depends on the type and the strength of the connection. Therefore, positive (negative) increment values are associated to excitatory (inhibitory) connections. Each neuron is implemented using a VHDL code in which the connection strength is selected to be a fraction of the neuron threshold (in particular, these fractions are selected to be $\pm 2/5$ and $\pm 1/10$).

2.2 Self-learning Architecture

The proposed self-learning architecture is shown in Fig. 2. It consists in two basic blocks, a Genetic Algorithm Circuitry (GAC) and a Fitness Circuitry (FC). The GAC generates new configurations based on the better configuration obtained, that is stored

in the configuration register. Using a Random Number Generator (RNG) a random mutation vector is generated. The mutation vector is operated using XOR gates with the better configuration found until the moment (placed in the configuration register). The result is a new configuration (binary output of XOR block) that is equal to the previous except in those cases where the RNG provides a HIGH state. The new configuration is directly applied to the SNN when the controlling signal of the GAC multiplexer (SL) is HIGH (self-learning selection). When signal SL is LOW (operation mode) the better configuration obtained until that moment is applied to the SNN.

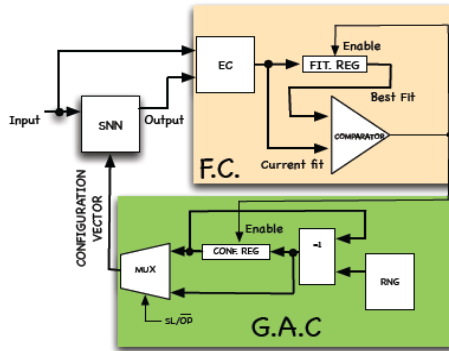


Fig. 2. Block structure for the dynamic configuration of SNN. The GAC block is used for the network configuration while the FC block evaluates the network efficiency.

The FC block evaluates the aptness of each new configuration for a selected network task. During the training mode an evaluation circuitry (EC) compares the SNN behavior with respect to the expected behavior, thus evaluating a cost function (the configuration fitness). The value obtained in this process is then compared to the one associated to the better configuration at the moment (stored at the fitness register). When a better fitness is found at the end of the evaluation time, the digital comparator output is set to a HIGH state and both the fitness and the configuration registers are updated with the new values. When the system is in operation mode, the SNN configuration is fixed to the better solution obtained at the moment. A global reset is used to start with pre-selected initial conditions.

2.3 Random Vector Generation

The mutation vector is used to generate a new configuration that is equal to the previous one, except in those cases where the RNG provides a HIGH state. The election of the RNG is important since all the possible mutation vectors must have the same probability of being generated. Therefore, the percentage of mutation ranges between the 0% (mutation vector 00...0) and the 100% (mutation vector FF...F). Using this strategy we ensure the possibility of moving from a local minima to a deeper (and therefore better) minima. Make note that, since the system is directly implemented in hardware it can sweep millions of different configuration vectors per second, thus obtaining a good solution in a reasonable time (although, of course, the absolute minimum is not guaranteed).

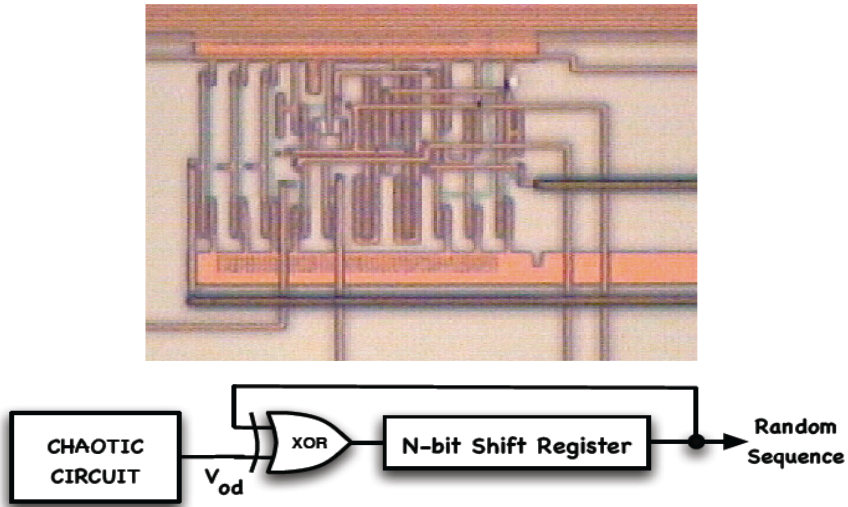


Fig. 3. Integrated circuits presenting a chaotic behavior can be used to quickly generate a mutation vector with N-bits. In this figure we can see an integrated prototype that present a chaotic behavior. The result is a completely random vector of N bits.

For the generation of the mutation vector we can choose either a pseudo-random or a random number generator. In FPGA applications we can use the first one since it is easily implemented using LFSR counters. For VLSI implementations we can choose a lower-cost solution as true random number generators [4]. The solution proposed in [4] represent a low cost in terms of hardware resources if compared to LFSR counters. The mutation vector is obtained using a chaotic circuit that generates a chaotic binary signal that is used to serially fill a shift register.

In Fig. 3 we show a photograph of an integrated circuit that is able to generate a chaotic binary signal. A N-bit shift register is serially filled with the product of performing the XOR function of the chaotic binary signal and the output bit of the register. In contrast to pseudo-random signal generators, the bit sequence generated does not present a periodic behavior (an LFSR of n-bits presents a period of repetition of 2^n-1 bits). For more details about the integrated circuit of Fig. 3 the reader is referred to the work in [4].

3 Application to Temporal Pattern Recognition

We applied the proposed SNN architecture to evaluate the processing behavior of various networks. The selected SNN task is the temporal pattern recognition (that is directly related to the “memory” capacity of the system). During the training mode, a finite sequence of vectors is repeatedly applied at the SNN input (the training bit sequence) and the task of the network consists in recognizing the sequence: at each time step the SNN has to provide the next bit of the sequence (see the illustration of Fig. 4). The network efficiency is evaluated estimating the probability of the SNN prediction success. At each evaluation step a mutated configuration provided by the GAC is used

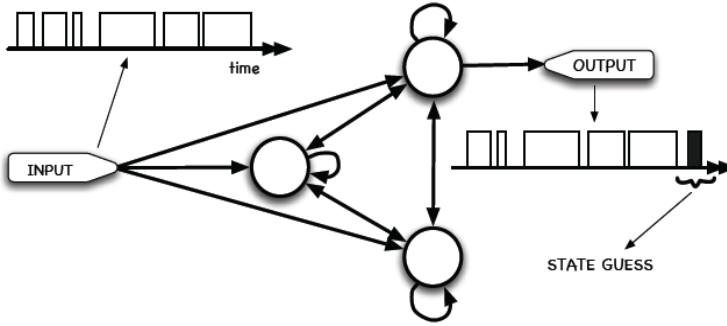


Fig. 4. A complete SNN implements all the possible inter-neuron connections. Such networks are trained to recognize temporal patterns.

to configure the SNN. The FC bloc evaluates the probability of success of SNN predictions and the mutated configuration is therefore stored or discarded (see Fig. 2).

We configured different networks containing three, four and five neurons, each one connected to the rest of the network thus assembling a complete topology. In Fig. 4 we show the case with three neurons (defined as a 3-SNN).

The training bit sequence must be as complex as possible to maximize the pattern recognition difficulty. Therefore, we selected the generation of pseudo-random strings provided by LFSR digital blocks. Pseudo-random bit strings are characterized to have the same statistical properties as random sequences with the only characteristic that pseudo-random sequences have a periodicity. In our experiment we used training sequences of N=7, 15, 31, 63, 127, 255, 511 and 1023 bits. With the selection of this type of pseudo-random sequences, the memorization task difficulty is maximized.

We applied each pseudo-random training sequence to three different SNN with complete topology (using 3, 4 and 5 neurons). Each SNN is configured by the proposed genetic-based self-learning architecture. At each time step, the network has to guess the next bit that will be provided by the LFSR. Once the configuring circuitry has been stabilized to an optimum configuration we evaluate the probability of success associated to this final configuration. In Table I we provide the different prediction success (in percentage) for each network and training sequence together with the results of the model explained next. It is observed that, as parameter N decreases and as the number of neurons increases the network has a higher prediction success.

From measurement data, we inferred a rule to estimate the prediction success of each network. This rule is found to provide very close results to the measured data as is shown in Table 1. The rule is an analytical expression that relates the success probability (p) to the training sequence length (N):

$$p = \frac{1}{2} + \sqrt{\frac{K}{N}} \tag{1}$$

where parameter K is dependent on the type of the network (number of neurons, connection topology, etc.). For the networks trained we obtained K=1.36, 1.68 and 2 for the 3-SNN, 4-SNN and 5-SNN networks respectively. The expression in Eq. (1), indicates that as the length of the pseudo-random sequence increases the success probability of the network decreases to the limit value of 0.5 (50% of success probability).

Table 1. Temporal recognition effectiveness of SNN (success probability)

N	3-SNN Model	4-SNN Model	5-SNN Model
7	86	94	100
15	80	80	83
31	71	71	73
63	67	65	66
127	62	60	62
255	59	57	58
511	56	55	56
1023	54	54	54

4 Conclusions

In this work we propose a simple architecture for SNN self-configuration. The proposed system implements in hardware a simplified genetic algorithm. We applied the proposed architecture to temporal pattern recognition analysis. Different pseudo-random sequences were applied to different networks and the success probability was evaluated. It is shown that prediction success seems to follow a law in which the success probability is inversely proportional to the square root of the number of bits of the sequence.

Acknowledgments. This work was supported in part by the Balearic Islands Government in part by the Regional European Development Funds (FEDER) and in part by the Spanish Government under projects PROGECIB-32A and TEC2008-04501.

References

1. Malaka, R., Buck, S.: Solving nonlinear optimization problems using networks of spiking neurons. In: Int. Joint Conf. on Neural Networks, Como, pp. 486–491 (2000)
2. Sala, D.M., Cios, K.J.: Solving graph algorithms with networks of spiking neurons. IEEE Trans. on Neural Net. 10, 953–957 (1999)
3. Gerstner, W., Kistler, W.M.: Spiking neuron models. Cambridge University Press, Cambridge (2002)
4. Rosselló, J.L., Canals, V., de Paúl, I., Bota, S., Morro, A.: A Simple CMOS Chaotic Integrated Circuit. IEICE Electronics Express 5, 1042–1048 (2008)

Research on Multi-Agent Parallel Computing Model of Hydrothermal Economic Dispatch in Power System

Bu-han Zhang¹, Junfang Li¹, Yan Li¹, Chengxiong Mao¹, Xin-bo Ruan¹,
and Jianhua Yang²

¹Electric Power Security and High Efficiency Lab, Huazhong University of
Science and Technology, Wuhan 430074, China

²Power Exchange Center of Central China Grid Company Limited,
State Grid of China, Wuhan 430074, China

Abstract. This paper presents a Multi-Agent Parallel Computing Model (MAPCM) for solving hydrothermal scheduling in the power system. The proposed model is hierarchical with two levels—decomposition level considering real time dispatch with optimal power flow and a coordination level considering the selection method of hydro-coal transfer coefficient under different situation about water head, and active power coordination with generation ramping constraints. The model is applied to a simple case study exploring the effect of multi-agent parallel computing about hydrothermal system. The efficiency analysis shows that the model has much potential for the real-time dispatching.

Keywords: Multi-agent parallel computing, Decomposition and coordination, Hydrothermal scheduling, Power system.

1 Introduction

Electricity industries world-wide are entering a period of implementing on-line energy manage system with precise and efficient calculation for daily dispatch, state analysis, and risk analysis. Electric power calculations are becoming more complex as grid sizes increase, especially considering electricity from renewable sources such as distributed hydropower station. Traditional economic dispatch problem in a bulk power system refers to active power dispatching of generation system, and can't consider both units operation constraints of active and reactive power and system constraints of transmission system, such as node voltage magnitude restricts and phase shift angle limits.

Optimal power flow (OPF) has developed for more than twenty years, and can accommodate types of controlled variable including active and reactive power injections, generator voltages, transformer tap ratios, and phase shift angles, even phase angle difference and node voltage magnitude. In a hydrothermal economic dispatch, the present OPF real-time off-line calculation are serial computation consuming a long time, and cannot satisfy the requirement of real-time on-line economic dispatch of large power system. It has become an urgent issue to develop efficient method for on-line dynamic hydrothermal real-time schedule. Parallel computation can cater for

the unending pursuit—ultrahigh speed. Multi-Agent based hydrothermal economic dispatching can solve the problem.

The multi-agent approach has been proposed in power system to solve the following issues: power system restoration[1-2], fault section and switching operation[3], electrical plant condition monitoring[4-5], test system for power system control[6], operation of distribution systems[7], outage work allocation[8], and control center infrastructure[9]. Besides the applications mentioned above, the multi-agent model can also be used in hydrothermal economic dispatch in power system.

A multi-agent system (MAS) is different from existing systems. It is a system comprising two or more agents or intelligent agents with three characteristics: reactivity, pro-activeness and social ability [10]. MAS should be considered for applications in an environment which display several characteristics [10]: a requirement for interaction between distinct conceptual entities, a very large number of entities in a system where it would be difficult to explicitly model overall system behavior, and enough data or information available locally to undertake an analysis or decision intelligently. In Energy Management System (EMS), which has the above characteristics, economic dispatching world-wide is facing a trend of energy conservation and sustainable development based on efficient and effective on-line computation platform. Multi-Agent Parallel Computing Model caters for this purpose.

This paper is concerned with the design and implementation of MAPCM in Economic Dispatching. There are three fundamental questions to be considered, namely:

- How should a physical model of MAPCM for economic dispatching be built?
- How should mathematical models of the computer simulation explain the physical model of MAPCM?
- How should a collection of agents communicate with each other and how to evaluate the efficiency?

This paper discusses a two-level model of intelligent agency—decomposition and coordination level. Consideration is given to coordinate hydrothermal economic dispatching and real-time communication of one agent with another agent entity in decomposition level. Finally, compared with the serial computation, standard of evaluation is given for MAPCM based economic dispatching.

2 Multi-Agent Parallel Computing Model

A multi-agent system consists of a collection of autonomous agents that communicate and collaborate to solve a complex problem [11]. In a power system, a multi-agent system offers organization plan by having agents as modular components that specialize some certain function of a composite problem. The MAPCM proposed in this paper differs from the Foundation for Intelligent Physical Agents' (FIPA) agent management reference model given by [12-13] aiming to define standards that can be used to support interoperability between agent-based systems.

2.1 Physical and Mathematical Modeling Approach Based on Two-level Agents

The physical model of MAPCM has two parts, the Master-agent and the distributed Agents according to time periods. In a 24 time periods real-time power scheduling per

day, shown in Fig. 1, Master-agent is the agent center that conserves data information and assigns real-time economic dispatching tasks to the Agents. Each node is named as Agent i that undergoes optimal power flow calculation in every time period i . The Agent exchanges information and communicates with the adjacent agent.

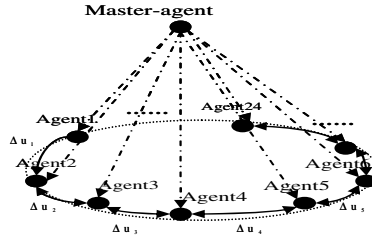


Fig. 1. Multi-agent Parallel Computing Model

The economic dispatching problem can be separated to two levels, decomposition level and coordination level. The mathematical theory of them is as follows.

1) **Decomposition level for real-time power dispatch**

The decomposition level is built for two purposes.

- For the Master-agent to assign the whole task of all time periods to every agent node i according to the assumptive time periods (eg.24 or 96).
- For the management of independent task of Master-agent and every other agent.

The hydrothermal economic dispatch is a complex optimization problem with inequality constraints for the control variables. It can be stated as

$$\min_{[P_{Tm}, P_{Hn} \in u_s]} \sum_{i \in T} F_i(x, P_{Tm}, P_{Hn}) \tag{1}$$

Subject to equality constraints in each time period i

$$\sum_{m \in R_{Ther}} P_{Tm} + \sum_{n \in R_{Hyd}} P_{Hn} - (P_{Di} + P_{Li}) = 0 \tag{2}$$

$$\sum_{n \in R_{Hyd}} W_n(P_{Hn}) = W_i \tag{3}$$

And subject to inequality constraints

$$P_{Tm}^{\min} \leq P_{Tm} \leq P_{Tm}^{\max}, P_{Hn}^{\min} \leq P_{Hn} \leq P_{Hn}^{\max} \tag{4-a}$$

$$Q_{Tm}^{\min} \leq Q_{Tm} \leq Q_{Tm}^{\max}, Q_{Hn}^{\min} \leq Q_{Hn} \leq Q_{Hn}^{\max}, V_{Tm}^{\min} \leq V_{Tm} \leq V_{Tm}^{\max},$$

$$V_{Hn}^{\min} \leq V_{Hn} \leq V_{Hn}^{\max}, \delta_{jk}^{\min} \leq \delta_{jk} \leq \delta_{jk}^{\max}, P_{tie} \leq P_l^{\max} \tag{4-b}$$

$$\Delta P_{Tm} \geq P_m^{Ra}, \sum_{m \in R_{ther}} \Delta P_{Tm} \geq P_i^{FIX} \tag{4-c}$$

Where,

- $F_i(x_i, u_i)$ is energy consumption function
- $[u_s]$ is a finite vector of the control variables
- $[x]$ is a vector of state variables x
- P_{Tm} is the active power of thermal generator m
- P_{Hn} is the active power of hydro generator n
- R_{ther} is a finite set of thermal generators
- R_{hyd} is a finite set of hydro generators
- P_{Di} is the load of the time period i
- P_{Li} is the transmission losses

- $W_n(P_{Hn})$ is the water consumption of each hydro generator n in the time period i
- W_i is the total water consumption

$[u^{min}]$, $[u^{max}]$ is the limits of the control variables vector including active power, reactive power, voltage magnitude and phase angle difference, namely,

$$u^{min} = [P_{Tm}^{min}, Q_{Tm}^{min}, V_{Tm}^{min}, P_{Hn}^{min}, Q_{Hn}^{min}, V_{Hn}^{min}, \delta_{jk}^{min}]^T, u^{max} = [P_{Tm}^{max}, Q_{Tm}^{max}, V_{Tm}^{max}, P_{Hn}^{max}, Q_{Hn}^{max}, V_{Hn}^{max}, \delta_{jk}^{max}]^T$$

- δ_{jk} is phase angle difference between node j and node k with the limits.
- P_{tie} is the active power of the tie line
- P_{tie}^{max} is the upper bound of the P_{tie}
- ΔP^{Tm} is the normal unit response rate (MW/hour) for unit m in hour i [14]
- P_m^{Ra} is the minimum ramp rate for unit m
- P_i^{FIX} is the required change in generation from hour $i-1$ to hour i [14]

Equation (2) is a system constraint about power balance. Equation (3) is the daily energy constraint of individual hydro unit. The two inequality constraints in (4-a) are the capacity constraints of individual thermal unit and hydro unit. The five inequality constraints in (4-b) are the bus state constraints. (4-c) includes two constraints. One is ramp rate constraint of the available thermal unit. The other is the system total ramp rate constraints.

This optimization problem is solved by introducing a vector of Lagrangian multipliers λ_i and minimizing the unconstrained Lagrangian function $L(x, u)$.

$$L(x, u) = \sum F_i(x, u) + [\lambda]^T [\sum_{m \in R_{Ther}} P_{Tm} + \sum_{n \in R_{Hyd}} P_{Hn} - (P_{Di} + P_{Li})] \tag{5}$$

Note that (5) contains objective function (1) and equality constraints (2) without considering (3). When (5) is satisfied, (3) is not satisfied in fact. As a matter of fact, (3) is an inequality condition, which has gradual convergence feature. It can be described exactly as

$$\max \left| \sum_{n \in R_{Hyd}} W_n(P_{Hn}) - W_i \right| < \varepsilon, \forall \varepsilon \rightarrow 0 \tag{6}$$

Assuming the energy consumption function is characterized by quadratic function, namely,

$$\begin{cases} F_i(P_{Tm}) = a_m + b_m P_{Tm} + P_{Tm}^2 \\ W_i(P_{Hn}) = a_n + b_n P_{Hn} + P_{Hn}^2 \end{cases}$$

According to the coordination equation approach, water consumption functions of n hydro units can be converted to n thermal energy consumption functions, as

$$\gamma_n \frac{dW_n(P_{Hn})}{dP_{Hn}} = \frac{dF_n(P_{Tn})}{dP_{Tn}}$$

where, γ_n is a hydro-coal transfer coefficient, much depends on the hydraulic head of the hydraulic generators. If the hydraulic head maintain the same, γ_n is a constant. Otherwise, γ_n changes with the hydraulic head. This is important for building this multi-agent model. Assuming the hydraulic head γ_n of each hydraulic generator is a continuous function indexed by time, namely $\gamma_n(t)$. In the MAPCM, shown in Fig. 2,

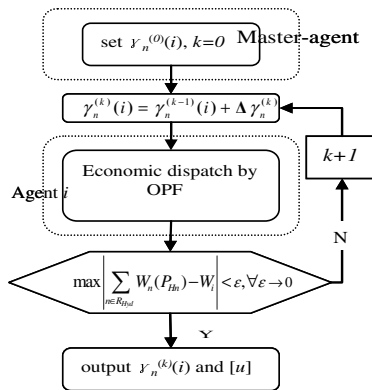


Fig. 2. Flow chart

the Master-agent can use an extra helper piecewise linear variable $\gamma_n(i)$ for each such $\gamma_n(t)$ in each time period t and set the initial value $\gamma_n^{(0)}(i)$. Agent i uses OPF method to solve the single objective optimization problem.

2) Coordination level for communication between agents

Coordination level is built for the following two communication behaviors:

- The interaction between the Master-agent and Agent i .
- The interaction between Agent i and Agent j .

As shown in Fig. 2, the Master-agent can calculate $\gamma_n(i)$ using an iteration method because it is difficult to work out $\gamma_n(i)$ in a complex and large scale composition hydrothermal system. The iteration method uses a steady incremental step $\Delta \gamma_n^{(k)}$, the value of which can be surplus or minus using the dichotomy method or the interpolation method according to the demand for precision.

In the time period i , a steady step $\Delta \gamma_n^{(k)}$ accumulate to $\gamma_n^{(k-1)}(i)$ step by step in order to form $\gamma_n^{(k)}(i)$, as

$$\gamma_n^{(k)}(i) = \gamma_n^{(k-1)}(i) + \Delta\gamma_n^{(k)}, k = 1, 2, 3, \dots, K$$

$$\Delta\gamma_n^{(k+1)} = \pm \frac{1}{2} \Delta\gamma_n^{(k)} \text{ or } \Delta\gamma_n^{(k+1)} = \frac{W_n^{(k+1)} - W_n^{(k)}}{W_n^{(k-1)} - W_n^{(k)}} \cdot [\gamma_n^{(k-1)}(i) - \gamma_n^{(k)}(i)]$$

where K is a upper bound of k , that satisfies the following inequality condition

$$\begin{cases} \sum_{n \in R_{Hyd}} W_n^{(0)}(P_{Hn}) - W_i > 0 \\ \sum_{n \in R_{Hyd}} W_n^{(K)}(P_{Hn}) - W_i < 0 \end{cases} \quad (7)$$

From inequalities (7), the iteration can stop only when $\gamma_n^{(k)}(i)$ satisfies

$$\sum_{n \in R_{Hyd}} W_n^{(k)}(P_{Hn}) - W_i = 0 \text{ or } \max_{i \in R_{Thd}} \left| \sum_{n \in R_{Thd}} W_n^{(k)}(P_{Hn}) - W_i \right| < \varepsilon, \forall \varepsilon \rightarrow 0 \quad (8)$$

2.2 Information Management

Information management aims to manage data that may be some physical and mathematical constraints. In a composite generation and transmission system, there are some operation constraints including the ramping limits of the thermal generators, and some system constraints for the control parameters and the state parameters, the latter including the maximum capacity of each tie line. These control variables which need to be interactive between the agents, can be ensured in this model.

The abstract architecture of the Master-agent and the Agents is presented in Table 1.

Table 1. Environment variables for master-agent

Variables	the Master-Agent	the Agents
[u]	V and δ of slack node, P and Q of each P,Q-node P and V of each P,V-node (read information)	the same as the Master-agent (deal with information)
[Δu]	none	[$\Delta P \Delta Q \Delta V \Delta \delta$] or only ΔP_T (coordinate level)
[x]	V and δ of each P,Q-node, δ of each P,V-node (read information)	the same as the Master-agent (deal with information)
Others	$\gamma_n^{(k)}(i)$ (coordinate level)	Cost to start up and shut down

Agents exchange accumulated vector [Δu_i] in order to restrict the active power increment ΔP_i and reactive power increment ΔQ_i of generator nodes, voltage magnitude increment ΔV_i and phase shift angle increment $\Delta \delta_i$ of generator nodes, namely,

$$[\Delta u_i] = [\Delta P_i, \Delta Q_i, \Delta V_i, \Delta \delta_i]^T$$

where,

$[\Delta u_i]$ is a control variable incremental vector.

In some cases, $[\Delta u_i]$ can't be considered, and use only ΔP_T instead. ΔP_T is the ramp rate constraint of the thermal generators. When the information about ΔP_T is delivered from the agent i to the agent $(i+1)$, the control variables $P_m(i+1)$ will be amended to become $P'_m(i+1)$ with regulated upper and lower bound, namely

$$\max\{P_m^{min}, P_m(i) - \Delta P_{Tm}\} \leq P'_m(i+1) \leq \min\{P_m^{max}, P_m(i) + \Delta P_{Tm}\}$$

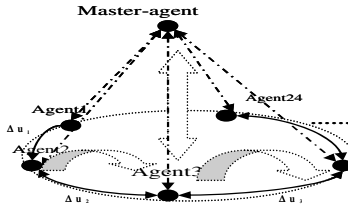


Fig. 3. Information flow in MAPCM

As shown in Fig. 3, in MAPCM, each agent has a unique ID number. Like the developed MAS, the agent with the smallest ID number is defined as the root agent of the system, which initiates the information flow in the system [15]. The Master-agent is the root agent of MAPCM. In the decomposition level, the Master-agent charges the independent task about reading information, while the Agents majors in dealing with information and reading information about cost of starting up and shutting down. In the coordination level, the Master-agent coordinates $\gamma_n^{(k)}(i)$ and transforms it to the Agents i , while the Agents coordinate information about $[\Delta u_i]$ and ramp rate ΔP_{Tm} .

3 Index for Efficiency Evaluation

For the purpose of the efficiency evaluation, the proposed indices of a MAPCM based composite generation and transmission systems depend on two basic aspects: the speed of calculation coordinating the Master-agent with the Agents and the time consumption of the repeated iteration computation in all of the Agents with the OPF. The latter aspect accounts for more time. With the scale of the power system increasing, the obvious advantage of the MAPCM can be seen.

(1) Absolute Cumulative Time (second) — ACT

ACT means the total time that the computation spends on the process from reading data information to printing the results. In a parallel computation, assuming each agent i spend $t_{agent}(i)$ in receiving information from Master-agent and executing the dispatch task in the MAPCM, ACT means the maximum value of $t_{agent}(i)$ and can be described as

$$ACT = \max_{1 \leq i \leq T} \{t_{agent}(i)\} \tag{9}$$

If repeating this work by a serial computation, absolute cumulative time is the sum of the time that each agent spends:

$$ACT^* = \sum_{i=1}^T t_{agent}(i) \tag{10}$$

Compared (9) with (10), it is obvious that the value of ACT obtained from (9) is rather smaller than that obtained from (10). In fact, $t_{agent}(i)$ is the sum of the time $t_a^{(j)}(i)$ that each iteration spend on OPF computation in order to find the appropriate $\gamma_n^{(k)}(i)$ to satisfy (8). Equation (9) can be stated exactly and completely as

$$ACT = \min_{1 \leq i \leq T} \max_{j=0}^k \{ t_a^{(j)}(i) \}, \text{ and } \gamma_n^{(k)}(i) \text{ satisfies (8)}$$

(2)Relative Time Saving Rate — RTSR

$$RTSR = \frac{ACT^* - ACT}{ACT^*}$$

To evaluate the effect of multi-agent parallel computation, RTSR indicates the relative efficiency by this approach compared with using a serial computation. The value of RTSR limits to an open interval (0, 1). The more the value of RTSR is, the more efficient this parallel computation can be. RTSR is affected by the number of the agents. The situation that one agent is only for one task has a high efficiency, while in practice the number of the agents is always pre-determined. The efficiency with the limited agents is lower than that with the unlimited ones.

4 Efficiency Analysis

Based on the above efficiency evaluation method, a test example has been taken for calculating the indices of IEEE 14-, 30-, 57-, and 118-bus test system in 12 hours and 24 hours using the following three assumptions:

- (1)Assume that the number of the agents is 24.
- (2)The original data information of bus, generator and branch is under the situation that the load ratio is 1.0 for the 18th time period. The load on each bus in every other time period is proportionally converted into the new case for each time period.
- (3)All of the computations converge after three-time iterations, namely $K=3$.

The simulation study of this test example was conducted on a AMD 64 Dual 5200+ computer. Though the process of finding right $\gamma_n^{(k)}(i)$ is added to the whole access, it spend a little time relatively compared with the OPF computation. The ACT is decided primarily by the time of the OPF computation. Table 2 shows the load ratio of each time period for 24 hours. The test example indices for these test system are given in Table 3.

Fig. 4 shows the $t_{agent}(i)$ in 24 time periods for 14-, 30-, 57- and 118-bus system. It is noted that the time increases approximately in the form of the exponential function with the scale of the power system increasing. RTSR given in Fig.5 has two characteristics. (i) In the same test system, the longer the time is, the more the value of RTSR can be. The slope of the curve is the sensitivity of RTSR with time periods.

(ii) In the same time period, the larger the scale of the test system, the more time can be saved. It can be concluded that the advantage of multi-agent parallel computation can be rather obvious with the increasing in the scale of power grid and the time when the number of the agents is equal to that of the time periods.

Table 2. Load ratios for 24 time periods

Time Period	Load Ratio	Time Period	Load Ratio
1	0.941	13	0.953
2	0.941	14	0.956
3	0.939	15	0.956
4	0.937	16	0.96
5	0.935	17	0.97
6	0.940	18	1.000
7	0.955	19	0.993
8	0.958	20	0.991
9	0.960	21	0.995
10	0.965	22	0.99
11	0.965	23	0.98
12	0.958	24	0.946

Table 3. Indices for test example

prog /CPU time	14-bus	30-bus	57-bus	118-bus
ACT- 24 t. (sec.)	1.76	7.56	13.64	4917.44
ACT-12 t. (sec.)	1.76	7.56	12.68	4917.44
ACT*-24 t (sec.)	37.96	129.56	284.48	94326.32
ACT*-12 t (sec.)	18.96	66.08	141.04	47142.40
RTSR-24 t	0.9536	0.9416	0.9521	0.9479
RTSR-12 t	0.9072	0.8856	0.9101	0.8957

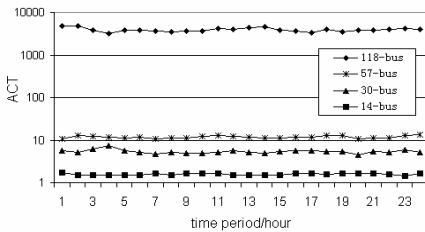


Fig. 4. The $t_{agent}(i)$ for IEEE 14-, 30-, 57-, 118-bus system on log plot

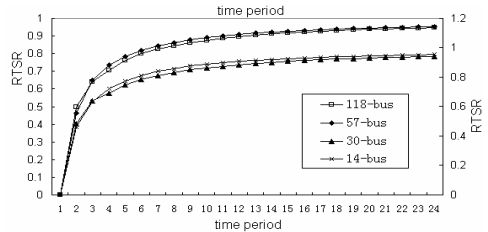


Fig. 5. RTSR in all time periods for 14-, 30-, 57, 118-bus system on bi-coordinate plot

It can be seen that RTSR for 24 time periods is around 0.95, indicating that multi-agent approach can save approximate 95 percent of working time spent by the serial computation. In practice, however, this advantage in efficiency can seldom be seen because this high efficiency value is obtained with the assumption that the number of iteration is three-time and the number of the agents is 24. When $\gamma_n^{(0)}(i)$ is chosen improperly, it is impossible that each iteration process runs the same three times as others in the economic dispatch of hydrothermal system. However, the cumulative CPU time is determined by the frequency of the iterations. If the frequency of the iteration in each ACT can be decreased, namely, the time spent on finding the appropriate $\gamma_n^{(k)}(i)$ be less, the RTSR can be increased. In the real-time dispatching, according to the historical data about $\gamma_n^{(k)}(i)$, $\gamma_n^{(0)}(i)$ can be chosen appropriately by the approximate evaluation, which makes the iteration run three or four times at most to find the appropriate $\gamma_n^{(k)}(i)$. RTSR can also be decreased by reduction in the agents.

The above analysis is based on assumption and reasoning. The proposed model can also consider the tie line constraints and several unit operating constraints such as minimum up and down times in a composite generation and transmission system. In a large scale hydrothermal power system, to consider all kinds of demands such as the adequacy and economical efficiency, the proposed economical dispatch approach based on Multi-agent Parallel Computing Model has tremendous potential for future development.

5 Conclusions

This paper presents a Multi-Agent Parallel Computing Model for hydrothermal economic dispatch in power system. The proposed model is hierarchical with two levels—decomposition level considering real time dispatch with optimal power flow and a coordination level considering the selection method of hydro-coal transfer coefficient under different situation about water head, and active power coordination with generation ramping constraints.

Case studies with IEEE 14-, 30-, 57-, and 118-bus system in 12 hours and 24 hours with three assumptions. Table 3 provides the comparable values for those indices proposed in part III. The efficiency of this proposed model is obvious. The model presented is a useful design for application of Multi-Agent System in the economic dispatch of the large scale hydrothermal power system, especially for nowadays policy to save energy, lower energy consumption and reduce pollutants discharge.

Acknowledgement. This work was supported in part by the Key Project of National Natural Science Foundation of China (50837003) and in part by the 973 Program (2009CB219702).

References

1. Nagata, T., Sasaki, H.: A Multi-Agent Approach to Power System Restoration. *IEEE Trans. on Power Systems* 17, 457–462 (2002)
2. Nagata, T., Tao, Y., Fujita, H.: An Autonomous Agent for Power System Restoration. In: *IEEE Power Engineering Society General Meeting*, pp. 1069–1074. IEEE Press, New York (2004)
3. Zhou, M., Ren, J., Li, G., Xu, X.: A Multi-Agent Based Dispatching Operation Instructing System in Electric Power Systems. In: *IEEE Power Engineering Society General Meeting*, pp. 436–440. IEEE Press, New York (2003)
4. Mangina, E.E., McArthur, S.D.J., McDonald, J.R.: The use of a Multi-Agent Paradigm in Electrical Plant Condition Monitoring. In: *IEEE Large Engineering Syst. Conference on Power Engineering (LESCOPE 2001)*, pp. 31–36. IEEE Press, New York (2001)
5. Catterson, V.M., Davidson, E.M., McArthur, S.D.J.: Issues in Integrating Existing Multi-agent Systems for Power Engineering Applications. In: *13th IEEE International Conference on Intelligent Systems Application to Power Systems*, pp. 396–401. IEEE Press, New York (2005)
6. Gehrke, O., Bindner, H.: Building a Test Platform for Agents in Power System Control: Experience from SYSLAB. In: *IEEE International Conference on Intelligent Systems Applications to Power Systems (ISAP 2007)*, pp. 1–5. IEEE Press, New York (2007)
7. Al-Hinai, A., Feliachi, A.: Application of Intelligent Control Agents in Power Systems with Distributed Generators. In: *IEEE Power Systems Conference and Exposition*, pp. 1514–1519. IEEE Press, New York (2004)
8. Zhao, N., Kawahara, K., Kubokawa, J., et al.: A Study of Outage Works Allocation by Means of Agent Technology. In: *IEEE International Conference on Power System Technology*, pp. 369–373. IEEE Press, New York (2000)
9. de Azevedo, G.P., Feijo, B.: Agents in Power System Control Centers. In: *IEEE Power Engineering Society General Meeting*, pp. 1040–1041. IEEE Press, New York (2005)

10. McArthur, S.D.J., Davidson, E.M., Catterson, V.M., Dimeas, A.L., Hatziargyriou, N.D., Ponci, F., Funabashi, T.: Multi-Agent Systems for Power Engineering Applications—Part I: Concepts, Approaches, and Technical Challenges. *IEEE Trans. on Power Systems* 22, 1743–1752 (2007)
11. Celaya, J.R., Desrochers, A.A., Graves, R.J.: Modeling and Analysis of Multi-Agent Systems Using Petri Nets. In: *IEEE International Conference on Systems, Man and Cybernetics*, pp. 1439–1444. IEEE Press, New York (2007)
12. McArthur, S.D.J., Davidson, E.M., Catterson, V.M.: Building Multi-agent Systems for Power Engineering Applications. In: *IEEE Power Engineering Society General Meeting (PES 2006)*, pp. 1–7. IEEE Press, New York (2006)
13. McArthur, S.D.J., Davidson, E.M., Catterson, V.M., Dimeas, A.L., Hatziargyriou, N.D., Ponci, F., Funabashi, T.: Multi-Agent Systems for Power Engineering Applications—Part II: Technologies, Standards, and Tools for Building Multi-agent Systems. *IEEE Trans. on Power Systems* 22, 1753–1759 (2007)
14. Peterson, W.L., Brammer, S.R.: A Capacity Based Lagrangian Relaxation Unit Commitment with Ramp Rate Constraints. *IEEE Transactions on Power Systems* 22, 1077–1084 (1995)
15. Huang, K., Srivastava, S.K., Cartes, D.A.: Solving the Information Accumulation Problem in Mesh Structured Agent System. *IEEE Transactions on Power Systems* 22, 493–495 (2007)

Fast Decoupled Power Flow Using Interval Arithmetic Considering Uncertainty in Power Systems

Shouxiang Wang, Chengshan Wang, Gaolei Zhang, and Ge Zhao

School of Electrical Engineering and Automation, Tianjin University,
Nankai District, Tianjin 300072, China

Abstract. In order to overcome the weakness of the conventional crisp analysis method, interval arithmetic is used in this paper to represent uncertainty in power systems. An interval fast decoupled power flow algorithm, which can be applied in large-scale power system analysis under uncertainty, is proposed. This proposed interval algorithm, taking many kinds of uncertainty into account, uses the Gauss elimination to solve two sets of interval linear equations, i.e. the decoupled P equations and Q equations. And the bounds of power flow results under uncertainty, namely the solution of interval flow equations, are obtained. Monte Carlo method based on fast decoupled power flow is also implemented to verify the validity of the proposed interval algorithm. The proposed method is testified with IEEE 14-node test system and a real large scale power system in North China. The test results illustrate the effectiveness and practical value of the proposed method by comparing with the results of Monte Carlo simulation and traditional crisp method.

Keywords: Uncertainty, power flow, Interval arithmetic, Monte Carlo simulation.

1 Introduction

More and more attention is paid to the problem and solution methods of uncertainty in the planning and operation of power systems. As a basic tool to power system analysis, power flow calculation encounters the difficulty in dealing with uncertainty. At present, there are three kind of algorithms of power flow calculation considering uncertainty in power systems: 1) fuzzy power flow algorithm [1], using fuzzy numbers to express uncertainty and calculate on the basis of fuzzy theory; 2) probability power flow algorithm [2-3], using probability theory to deal with the uncertainty; 3) interval algorithm [4-5], using interval number and interval arithmetic to process uncertainty which has explicit extension and indefinite connotation, or the information whose bounds are fixed and known exactly while whose accurate value is uncertain.

Interval Arithmetic [6] was established by R.E.Moore et al in 1960s, and has been applied widely in fields such as physics, chemistry, engineering, social science, etc. In engineering field, when we don't know the primary data of a problem exactly while we know that they are included in given ranges instead. In other words, the primary data are intervals instead of crisp point values, in such condition interval arithmetic can be

used to obtain the interval solution or the range over which solution of the present problems covers.

It is not a problem now to apply interval arithmetic into power flow calculation of large-scale distribution system, since the distribution network is mostly of radial structure and interval arithmetic can be easily introduced into the backward-forward sweep power flow algorithm, which is widely used in radial distribution network.

High-voltage transmission network however is generally of the loop structure, its power flow calculation problem in mathematics is ascribed to solve a set of multi-dimensional nonlinear equations, and the direct substitution of general mathematical operation with interval operation is not feasible. It is quite difficult to apply interval arithmetic into large-scale transmission systems. Zian Wang and Fernando L. Alvarado present an early discussion of interval arithmetic in power flow analysis [4]. They used interval arithmetic to solve power flow problem with interval input data. To solve interval non-linear equations, the Newton operator was used. For each iteration, the Gauss-Seidel Method was used to solve the interval linear equations. But the difficulty of application of interval arithmetic in large-scale power system was not solved in their paper. The convergence of interval power flow is not good and the result appeared too conservative which was far away from the true bound. In other literatures, other interval iterative methods such as Krawczyk-Moore method are also used. The common shortcomings of general interval iteration methods are that they are complex and need large amount of computation, they are not suitable for the solution of large-scale problems.

This paper proposes an interval fast decoupled power flow algorithm under uncertainty, which introduces interval arithmetic into traditional fast decoupled power flow [7] widely used in power systems. In this proposed algorithm, interval numbers are used to express the uncertainty of load variety. And the interval extension model of fast decoupled power flow is built. Then it is ascribed to the solution of two sets of interval equations with P, Q decoupled and interval Gauss elimination is applied to solve each set of interval linear equations independently. Through the iteration of P, Q equations separately, the interval computation result is obtained. For the purpose of comparison and verification, Monte Carlo simulation based on fast decoupled power flow considering uncertainty is also implemented. Test results verify the validity and value of the proposed interval power flow algorithm.

2 Interval Arithmetic and Interval Equations Solution

An interval number is defined as a pair of real numbers

$$[X] = [\underline{x}, \bar{x}] = \{x \in R \mid \underline{x} \leq x \leq \bar{x}\}, \quad (1)$$

where $\underline{x}, \bar{x} \in R$, $\underline{x} \leq \bar{x}$, and \underline{x}, \bar{x} are known as the inferior bound and superior bound of the interval number $[X]$, respectively. A rational number k is represented as an interval number $[k, k]$.

Let $[X] = [\underline{x}, \bar{x}]$ and $[Y] = [\underline{y}, \bar{y}]$ be the two interval numbers, then addition, subtraction, multiplication and division of these two interval numbers are defined as below:

$$[X] + [Y] = [\underline{x} + \underline{y}, \bar{x} + \bar{y}], \tag{2}$$

$$[X] - [Y] = [\underline{x} - \bar{y}, \bar{x} - \underline{y}], \tag{3}$$

$$[X] \cdot [Y] = [\min(\underline{x}\underline{y}, \underline{x}\bar{y}, \bar{x}\underline{y}, \bar{x}\bar{y}), \max(\underline{x}\underline{y}, \underline{x}\bar{y}, \bar{x}\underline{y}, \bar{x}\bar{y})], \tag{4}$$

$$[X]/[Y] = [\underline{x}, \bar{x}] \cdot [1/\bar{y}, 1/\underline{y}], \quad \text{if } 0 \notin [Y]. \tag{5}$$

Consider a set of interval linear equations

$$[A][X] = [B], \tag{6}$$

where $[A]$ is an interval coefficient matrix ($[A] \in I(R^{n \times n})$),

$[X]$ is an interval solution vector ($[X] \in I(R^{n \times 1})$),

$[B]$ is an interval constant vector ($[B] \in I(R^{n \times 1})$).

Equation (6) is the interval expansion of linear equations with fixed parameters to interval linear equations with interval parameters considering uncertainty.

The solution of interval linear equations is very different from that of the ordinary linear equations. Many researches have been done on solving interval linear equations [8-9], among which interval Gauss elimination method is widely used. In this paper, interval Gauss elimination is adopted too to solve interval linear equations in fast decoupled power flow. The procedure of this method consists of three steps as below.

1) Elimination process

For some $k < n$, making the following three operations from 1 to $n-1$:

(1) Choose the biggest element in absolute value from the right bottom sub-matrix starting from the k th line, the k th row of coefficient matrix A , and exchange it to the pivot position by row and column exchanges.

(2) Normalization

$$a_{kj} / a_{kk} \Rightarrow a_{kj}, \quad j = k + 1, \dots, n \tag{7}$$

$$[b_k] / a_{kk} \Rightarrow [b_k]. \tag{8}$$

(3) Elimination

$$a_{ij} - a_{ik}a_{kj} \Rightarrow a_{ij}, \quad j = k + 1, \dots, n \tag{9}$$

$$[b_i] - a_{ik}[b_k] \Rightarrow [b_i], \quad i = k + 1, \dots, n. \tag{10}$$

2) Back substitution process

$$(1) \quad [b_n] / a_{nn} \Rightarrow [x_n] \tag{11}$$

$$(2) \quad [b_i] - \sum_{j=i+1}^n a_{ij}[x_j] \Rightarrow [x_i], \quad i = n - 1, \dots, 1, 0. \tag{12}$$

3) Finally, adjust the order of the elements in the solution vector.

3 Fast Decoupled Power Flow under Uncertainty Using Interval Algorithm

The basic idea of the fast decoupled power flow using interval arithmetic being proposed in this paper is as follows. First, use interval numbers to express the uncertain variables in power system such as the uncertainty of load variety. Then, the mathematical model of the fast decoupled power flow using interval arithmetic is built as below:

$$[\Delta P] / [V] = [B'] [V] [\Delta \theta], \tag{13}$$

$$[\Delta Q] / [V] = [B''] [\Delta V]. \tag{14}$$

Note that in equation (13) and (14), the branch parameter matrixes are interval matrixes and state variable vectors are interval vectors, which mean that the elements of them are mostly interval numbers.

In the operation of actual power system, the influence of parameter uncertainty of electric lines and transformers factor is often small enough to be neglected, so the equations (13) and (14) can be simplified as below.

$$[\Delta P] / [V] = B' [V] [\Delta \theta], \tag{15}$$

$$[\Delta Q] / [V] = B'' [\Delta V]. \tag{16}$$

After the interval model of the fast decoupled power flow is established, the solution of the mathematic model becomes essential. The essence of the interval fast decoupled power flow is the solution of two sets of interval linear equations corresponding to P and Q iterations respectively. It is also the difficulty of implementing interval fast decoupled power flow. The aim of the interval power flow is to find an interval solution of node voltages and the phase angles with a boundary as small as possible, which reflects the influence on power flow of the uncertainty of generator actual active and reactive output, and load variety.

This paper has implemented the proposed algorithm with C++ standard language. The program is based on a traditional fast decoupled power flow program. In the traditional power flow program, sparse matrix and sparse vector technique are adopted. So it can be used to solve large-scale power systems. And its accuracy has also been verified with test data in literatures.

Monte Carlo simulation based on traditional fast decoupled power flow considering uncertainty is also implemented in this paper to verify the accuracy and the validity of the result obtained from the proposed interval method. The key points of Monte Carlo simulation method based on fast decoupled power flow are as follows: 1) sample randomly in variety intervals of uncertain parameters; 2) calculate crisp value of system state using traditional fast decoupled power flow algorithm; 3) get the upper bound and the lower bound of the system state from random sampling.

4 Numerical Results

To illustrate the application of interval fast decoupled power flow, two systems of the IEEE 14-node standard test system (see fig. 1) and the North China Power Grid have been chosen.

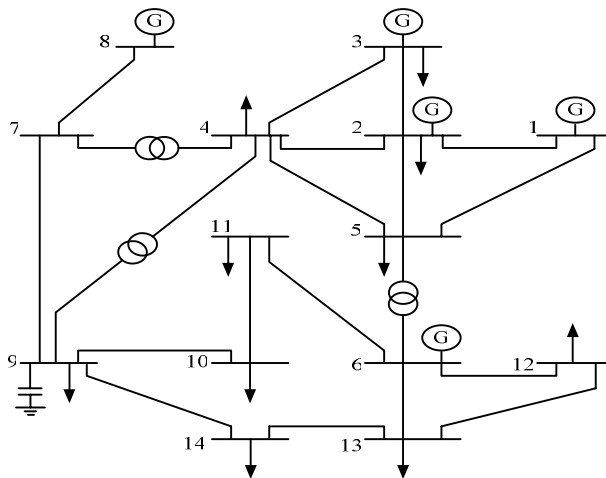


Fig. 1. IEEE 14-node test system

Taking the 14-node test system for example, two different cases with different degree of uncertainty are studied. In each case, the result of the fast decoupled power flow under certain conditions, the result of interval decoupled power flow under uncertainty, and the result of Monte Carlo simulation are listed, analyzed and compared.

Case 1. consider the uncertainty of load power only while power outputs of all generators are assumed fixed.

Table 1 and Fig. 2 show the result of the node voltage magnitude of the fast decoupled power flow under certain conditions, interval decoupled power flow under uncertainty, and Monte Carlo simulation (range of variation is $\pm 5\%$).

Table 1. Voltage magnitude (p.u.) (range of variation is $\pm 5\%$, only consider uncertainty of load power)

Node	Traditional crisp method	Under uncertainty	
		Interval method	Monte Carlo simulation
1	1.06000	[1.06000, 1.06000]	[1.06000, 1.06000]
2	1.04500	[1.04500, 1.04500]	[1.04500, 1.04500]
3	1.01000	[1.01000, 1.01000]	[1.01000, 1.01000]
4	1.01367	[0.99014, 1.03744]	[1.01220, 1.01509]
5	1.01635	[0.99720, 1.03615]	[1.01497, 1.01769]
6	1.07000	[1.07000, 1.07000]	[1.07000, 1.07000]
7	1.05008	[1.01386, 1.08230]	[1.04860, 1.05153]
8	1.09000	[1.09000, 1.09000]	[1.09000, 1.09000]
9	1.03344	[0.97086, 1.08911]	[1.03116, 1.03568]
10	1.03234	[0.96107, 1.09614]	[1.03006, 1.03457]
11	1.04737	[0.99826, 1.09487]	[1.04604, 1.04869]
12	1.05552	[1.02161, 1.10492]	[1.05475, 1.05628]
13	1.04727	[1.00633, 1.09110]	[1.04604, 1.04850]
14	1.02122	[0.94993, 1.08445]	[1.01842, 1.02399]

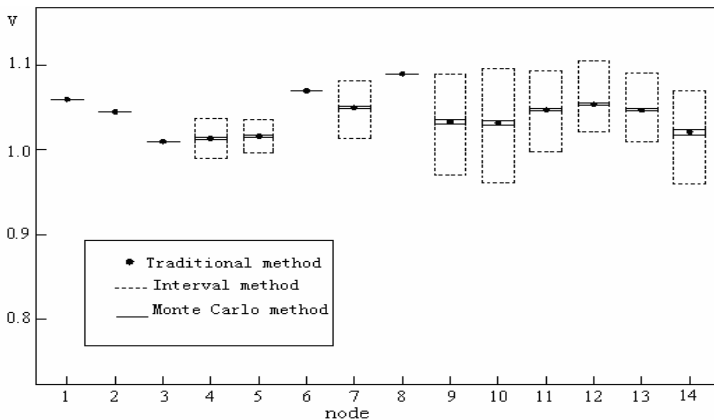


Fig. 2. Node voltage magnitude (p.u.) in case 1 (range of variation is $\pm 5\%$)

From Table 1 and Fig. 1, we can see that the result from interval method encloses the result of traditional crisp method when the parameter being set the midpoint of given interval, at the same time it also encloses the result calculated by Monte Carlo method. This demonstrates the validity of the proposed method. It verifies the completeness of interval arithmetic and interval method.

Result of the interval method under different degree of uncertainty is shown in Fig. 3 when the range of variation is $\pm 2\%$, $\pm 5\%$, $\pm 10\%$ respectively. It again shows that the result of interval method encloses the result of traditional crisp method and that of the Monte Carlo method. And it can be also found that the smaller the range of variation, the smaller the resulted interval, and the closer of the result of interval method to Monte Carlo method.

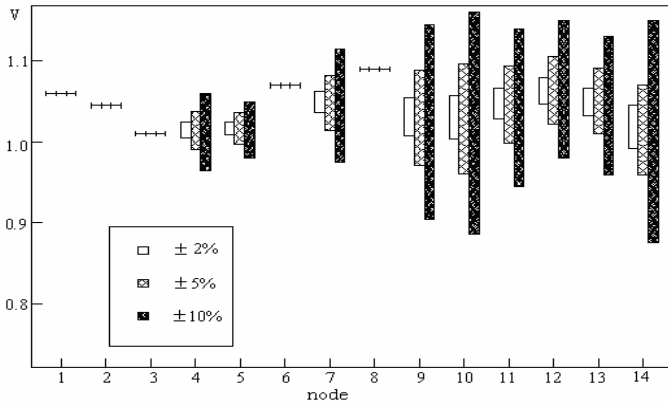


Fig. 3. Node voltage magnitude under different degree of uncertainty (p.u.)

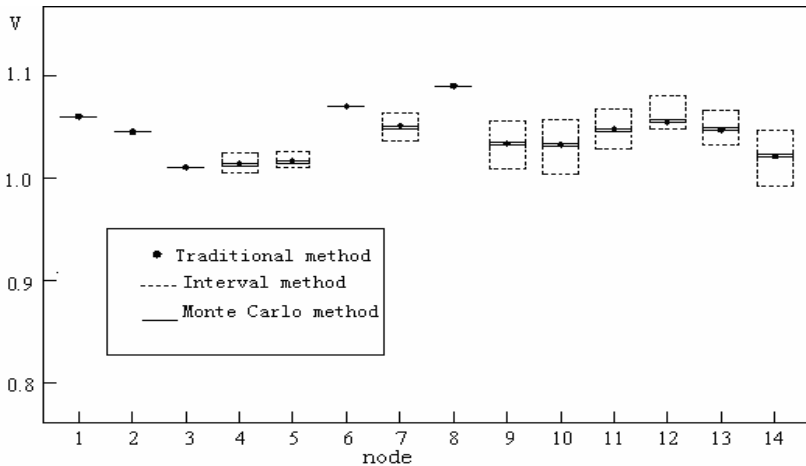


Fig. 4. Node voltage magnitude (p.u.) in case 2

Case 2. consider the uncertainty of load power and generator power.

Fig. 4 shows the result of the node voltage magnitude of the fast decoupled power flow under certain conditions, interval decoupled power flow under uncertainty, and Monte Carlo simulation (range of variation is $\pm 2\%$).

Through the above two case studies, the accuracy and validity of the proposed fast decoupled power flow using interval arithmetic have been verified. The proposed method is computationally superior to Monte Carlo simulation method although the result of the former is a little conservative than that of the latter.

The proposed interval power flow method can be effectively used in large-scale power systems such as the power grid of North China. The result is omitted here for space limitation.

5 Conclusion

Comparing to traditional crisp method, the proposed interval algorithm can take many kinds of uncertainty into account. And the bounds of power flow results under uncertainty, namely the solution of interval flow equations, are finally obtained. Hence operators in power system have got more information than before.

Comparing to Monte Carlo method, the proposed method is computationally superior to Monte Carlo simulation method although the result of the former is a little conservative than that of the latter.

The proposed method is testified with IEEE 14-node test system and a real large scale power system in North China. The test results illustrate the effectiveness and practical value of the proposed method.

References

1. Das, D., Ghosh, S., Srinivas, D.K.: Fuzzy Distribution Load Flow. *Electric Machines and Power Systems* 27, 1215–1226 (1999)
2. Dopazo, J.F., Klitin, O.A., Sasson, A.M.: Stochastic power flow method. *IEEE Trans* 94, 299–309 (1975)
3. Ding, M., Li, S.H.: Probabilistic Load Flow Analysis Based on Monte-Carlo Simulation. *Power System Technology* 25, 10–22 (2001)
4. Wang, Z., Alvarado, F.L.: Interval Arithmetic in Power Flow Analysis. *IEEE Trans on Power Systems* 7, 1341–1349 (1992)
5. Das, B.: Radial Distribution System Power Flow Using Interval Arithmetic. *Electrical Power and Energy Systems* 24, 827–836 (2002)
6. Moore, R.E.: *Interval Analysis*. Englewood Cliffs, New Jersey (1966)
7. Stott, B., Alsac, O.: Fast Decoupled Load Flow. *IEEE Trans on Power Apparatus and System PAS-93*, 859–869 (1974)
8. Neumaier, A.: New Techniques for the Analysis of Linear Interval Equations. *Linear Algebra Appl.* 58, 786–793 (1984)
9. Sudarsanam, A., Aravind, D.: A Fast and Efficient FPGA-based Implementation for Solving a System of Linear Interval Equations. In: 2005 IEEE International Conference on Field-Programmable Technology, pp. 291–292. IEEE Press, New York (2005)

Power System Aggregate Load Area Dynamic Modeling by Learning Based on WAMS

Huimin Yang and Jinyu Wen

Power Security and High Efficiency Lab,
Huazhong University of Science and Technology,
Wuhan 430074, China

Abstract. This paper is concerned with an investigation of a methodology using intelligent learning techniques based on WAMS to construct power system load area model. An aggregate load area dynamic model (ALADM) is proposed to represent large area loads of power system. A population diversity-based genetic algorithm (GA) combined with the recursive least squares (RLS) method is used to obtain the structure and parameters of the load model. Simulation results on EPRI 36-bus power system is given to show the potential of this new methodology of power system modeling.

Keywords: Genetic algorithm, Recursive least squares, Load area modeling, ALADM, Power system.

1 Introduction

There hasn't been any research about dynamic modeling for complex load centre of power system. Previous studies about load area modeling mostly concerning power system with simple structure, using bus load modeling method [1]. To modeling load centre with complex structure, the effective data measurement is prerequisite. But it is impossible to realize in the past. Recently, wide area measurement system (WAMS) and the corresponding information transmission network have been developed. The real-time information of different nodes in power system can be obtained by WAMS, which makes modeling of complex load area to be possible [2].

Ref [3] give a static modeling method for complex load centre. But it doesn't consider dynamic characteristics of power system components because of the complexity. So, besides data measurement, the complexity and excessive calculation in dynamic load modeling is another difficulty.

System parameter identification techniques have been widely studied for system modeling. The observed stimulus-response data are usually used to identify the model parameters. An error criterion of the response data is an objective function to be minimized, which is typically a function of the squared predictive errors. Least squares estimation (LS) is widely used in parameter identification for its unbiasedness and effectiveness. The recursive least squares (RLS) is a common method derived by LS and it is more efficient in estimation [4]. However, this quadratic mapping between the predictive errors and the identified model parameters is generally a complex,

nonlinear, and possibly nonconvex function in multi-parameters identification of complex system. LS could not deal with the problems of local minima, poor robustness and sensitivity to initial condition. It tends to diverge when used in real systems where uncertainties exist [5].

Genetic algorithms (GAs) [6] provides a new way to achieve system modeling, in particular for large-scale and complex industrial systems. It's one of the simulated evolution method which is a newly developed learning method. The technique is inspired by natural phenomena and simulated from natural genetics and evolution processes. It is population-based algorithms and provides a stochastic search capability which overcomes the shortcomings of LS [7]–[10]. However, to ensure identification accuracy of power system complex structures parameters, excessive calculation is a problem needs to be considered [11].

Combine the GA and RLS is an appropriate way. First, use GA to obtain values of parameters needed to be identified and the results can be used as the initial value of RLS. Then, they can be further optimized by RLS quickly through a little calculation. It can not only deal with the local minima and sensitivity to initial condition problem but also can satisfy high accuracy by less calculation.

This paper proposes power system aggregate load area dynamic model (ALADM) to represent load centre based on WAMS. The population diversity-based genetic algorithm (PDGA) and RLS are combined to identify the parameters of ALADM. Simulation results on EPRI 36-bus power system is given to show the potential of this new methodology of power system modeling.

2 Formulation of the Load Modeling

2.1 ALADM

There are some standard models recommended by IEEE for power flow and dynamic simulation programs. Most of these models are ‘bus load’ [12], [13], that is: a portion of the system that is not explicitly represented in a system model, but rather is treated as if it was a single power-consuming device connected to a bus in the system model. In practice, there is usually a complex load area to be tackled as an equivalent ‘area load’ as in Fig. 1(a). At this time, the power system can be classified into three parts: load area part, remainder system (or external system) part, and a connecting part (or boundary line) between these two sections. Under many conditions, this connecting part includes some main buses which may be the secondary buses of substations and/or the output buses of generating stations.

In Fig. 1(a), there are n buses between the load area and the external system. P_i and Q_i ($i=1, 2, \dots, n$) are the active and reactive power injected into the i th bus from external system. \dot{V}_i and \dot{f}_i are the i th bus voltage and frequency, respectively. The value of the above operating states can be measured by WAMS in an industrial power system. It should be pointed out here that the totally consumed active power

$P_l = \sum_{i=1}^n P_i$ and reactive power $Q_l = \sum_{i=1}^n Q_i$ (including the loss of the distribution network). It can be seen that the bus load is the special case ($n=1$) of the area load.

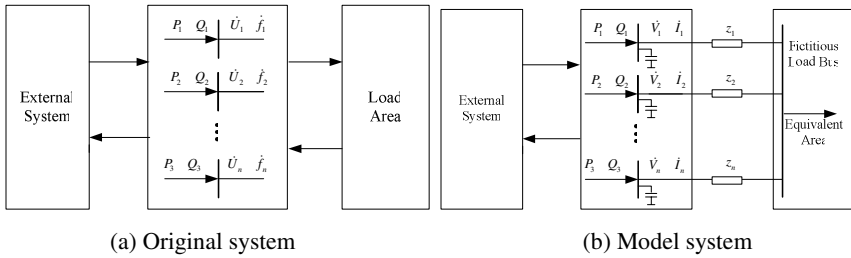


Fig. 1. Power system with a complex load area

As shown in Fig.1(b), ALADM is assumed to be connected to a fictitious load bus directly, and this load bus connects to each bus i through a fictitious transmission line. There are three parameters of the transmission line: the line resistance, the line reactance, and the shunt susceptance at the i th bus. The idea of the fictitious bus has been used for other applications such as the generator dynamic equivalence [14], [15].

ALADM should represent the relationship between the total power and the voltage and frequency of each bus with which the load area is connected. That means the model of the load area can be represented by the following equations:

$$\begin{cases} P_l = F_p(\dot{V}_1, \dot{V}_2, \dots, \dot{V}_n, f_1, f_2, \dots, f_n) \\ Q_l = F_q(\dot{V}_1, \dot{V}_2, \dots, \dot{V}_n, f_1, f_2, \dots, f_n) \end{cases} \quad (1)$$

Subject to the constraints

$$\begin{cases} P_{ic} = P_{im} \\ Q_{ic} = Q_{im} \\ \dot{V}_{ic} = \dot{V}_{im} \\ f_{ic} = f_{im} \end{cases} \quad (2)$$

where P_{im} is the measured active power injected into the i th bus in original system (Fig.1(a)), and P_{ic} is the calculated power in model system (Fig.1(b)). Other variables are similar to it.

2.2 Evaluation of Model Accuracy

When the comprehensive load model is used to model a load area, the connecting line parameters also need to be determined. An error between the data measured from the practical system and those computed from the model system given below can be used to evaluate the accuracy of the model system.

$$e(X) = \sum_{i=1}^n (k_p |P_{im} - P_{ic}| + k_q |Q_{im} - Q_{ic}| + k_v |\dot{V}_{im} - \dot{V}_{ic}| + k_f |f_{im} - f_{ic}|), \quad (3)$$

where $e(X)$ implies that the error is mainly affected by X which is a vector of the parameters to be estimated and consists of two parts, $X=[Z, Y]$: Z denotes a set of z_i ,

($i=1,2,\dots, n$) consists of the i th line resistance, reactance, and the shunt susceptance at the i th bus; Y denotes the coefficients of (1), k_p, k_q, k_r and k_f are the weighting factors. Noting that (3) is the system error at a time instant, a series of measured data has to be used to get the precise parameter values; therefore, the practically used error criterion is

$$E(X) = \sum_{i=1}^N e(X), \tag{4}$$

where i denotes time instant and N is the total number of samples. The load modeling process is to find a set of parameters which minimizes $E(X)$. It is a parameter optimization problem. As described in the introduction, a PDGA-RLS method is used to solve this problem.

3 GA-Based Load Model Parameter Identification

3.1 PDGA

Basic GA is often binary coded in bits (i.e., 0 or 1 s) and concatenated as a string which refers to a point in the solution space. A disadvantage of the binary-coded method is that when using GA for a multi-parameter problem, the length of each individual string has to be very long in order to get a sufficiently precise solution, but as the length increases, the efficiency of the GA reduces and the population evolves more slowly. On the other hand, if the length of the string is reduced, the solution precision may not be satisfactory.

In PDGA, the length of each string can be much shorter than a conventional GA, but a high precision solution can still be obtained fairly quickly. In this research, population diversity is introduced to solve this problem. In fact, the population diversity described before provides important information about the global optimal solution. As GA is a global search method, the population converges toward the global optimal solution. In a multi-parameter problem, the GA cannot find the best solution when using a short string length, but the last population should have covered the global optimal solution area. Obviously, the smaller the population diversity is, the smaller the area will be in which the global optimal solution can be found. Assume that the low precision solution when the GA stalls is $x_1^0, x_2^0, \dots, x_n^0$, the population diversity is $D(P^0)$ at this time. The new search space (a_k^0, b_k^0) for the k th parameter can be set to

$$\begin{cases} a_k^0 = \max[x_k^0 - D(P^0) \times |x_k^0|, a_k] \\ b_k^0 = \max[x_k^0 - D(P^0) \times |x_k^0|, b_k] \end{cases} \tag{5}$$

Obviously, the new search space is smaller than the previous one. Therefore, in the new space, a string with the same length codes parameters with higher precision. The PDGA then, regenerates a new population with the same string length and continues to search for the best solution in the new space. This processes each time when the PDGA stalls until the set convergence is achieved.

The detailed process of application of PDGA to an optimization problem can be found in [16].

3.2 PDGA-RLS Based Parameter Identification

When using PDGA to optimize the parameters of the ALADM, expensive computation will be involved in the search process, as the power system load flow calculation will be called for each fitness computation. To simplify the process, Z and Y in X are decoupled as follows to avoid calling load flow calculation each time, which significantly saves computation time. A two-step optimization procedure is proposed. The PDGA is employed to optimize the fictitious connecting line impedances first and use RLS for further optimization. Then used again for the load model coefficients.

Referring to Fig. 1(b), the voltage value of the fictitious load bus \dot{V}_s can be calculated out according to the following equation:

$$\dot{V}_s^i = \dot{V}_i - \dot{I}_i z_i, \tag{6}$$

where \dot{V}_i and \dot{I}_i are the measured voltage and current values at the i th connecting bus; z_i is the impedance value of the i th connecting line as defined before. Since there are n connecting lines, n voltage values for the fictitious load bus might be calculated and they should be equal to each other. As these z_i are the parameters to be optimized, the differences between these voltage values $\dot{V}_s^i (i=1,2,\dots,n)$ can be used to measure the quality of Z , that is

$$e(Z) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \left| \dot{V}_s^i - \dot{V}_s^j \right|, \tag{7}$$

where Z can be optimized from (6) and (7). Then, \dot{V}_s can be calculated out according to the optimal Z . After that, the power consumption in the fictitious load bus P_s and Q_s can be calculated by the following equation:

$$\begin{cases} P_s = \sum_{i=1}^n V_s I_i \cos \theta_i \\ Q_s = \sum_{i=1}^n V_s I_i \sin \theta_i \end{cases}, \tag{8}$$

where I_i is the measured value of the current flowing into the i th bus in Fig. 1(b); θ_i is the phase angle between V_s and I_i . With the \dot{V}_s obtained, corresponding P_l and Q_l for the load model at this bus can be calculated from (1). P_l and Q_l should be equal to P_s and Q_s , respectively. Therefore, the differences between P_l , Q_l and P_s , Q_s can be used to evaluate the quality of the coefficients, that is

$$e(Y) = k_p |P_s - P_l| + k_q |Q_s - Q_l|, \tag{9}$$

Y can be optimized from (8) and (9).

The overall procedure of the parameter optimization using the PDGA for the comprehensive load model is realized in the following steps.

- Step 1) PDGA is applied to optimize the impedance values of the n fictitious connecting lines, using (6), (7), and (4), then use RLS to optimize the coefficients further;
- Step 2) upon the optimization of the impedances, the PDGA is employed again to find the optimal coefficients in (1) based on (8), (9), and (4), then use RLS to optimize the coefficients further.

4 Simulation Results

The proposed method has been used to construct area load models for EPRI 36-bus power system shown in Fig. 3. This system consists of 8 units and 9 loads. There are two power supply area shown in dotted rectangle consists of 5 units. The load centre shown in the rectangle consists of 6 loads on bus 16, bus 18, bus 19, bus 20, bus 21 and bus 29 separately. There are also 3 small units in load centre. About 1000 MW power is transmitted to the load centre via 5 transmission lines from the two power supply areas. The data of the units, the main transformers, and the original network in the load area are given in [17].

Units are represented by five order model considering excitation system and governor. Loads in original system and equivalent with ALADM are all represented by dynamic load model of induction motor model paralleled with a constant impedance shown in Fig.2. The induction motor model is in the form shown in [16]. And the load coefficients of ALADM to be determined are: $[R_2, X_2, X_l, \alpha, P, s(0), T_{dob}, T_{jL}, K_{pq}], K_{pq}$ is the ratio of the constant impedance load in total load, others are the parameter of induction motor[17].

The load area is linked to the external system by five lines. When the performance of the external system is concerned, this area can be tackled as an area load. The ALADM can be used to represent this load area and the model system shown in Fig. 4 can be obtained.

Now the modeling process is regarded as determining the coefficients of load and the impedance values of 5 fictitious lines to minimize (9) and (7). Set three-phase short circuit fault on line between bus 22 and bus 23 and system dynamic response can be obtained. The PDGA-RLS is then used to identify the above coefficients. Results are shown in Table 1 and Table 2.

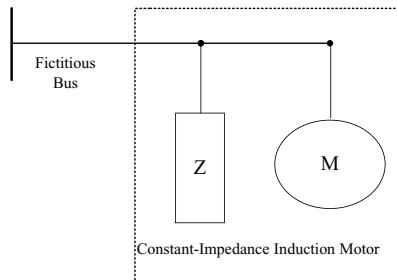


Fig. 2. Structure of load model

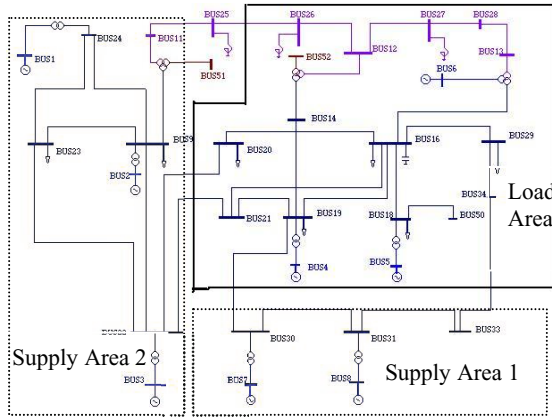


Fig. 3. 36-Bus power system with a comprehensive load model

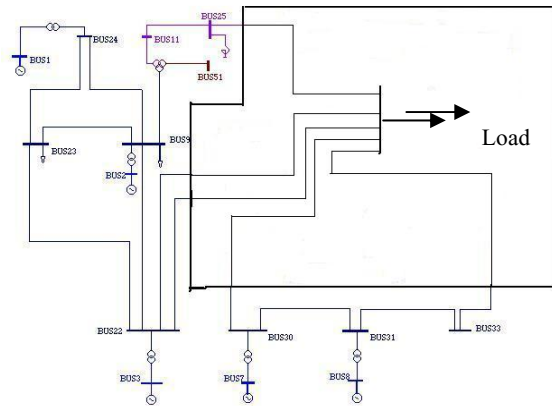


Fig. 4. Model power system with ALADM

Table 1. Parameters of fictitious Connecting lines

Line	R(p.u.)	X(p.u.)	B/2(p.u.)
1	0.0307	0.0553	0.4000
2	0.0312	0.0054	0.4000
3	0.0010	0.0600	0.4000
4	0.0070	0.0308	0.0200
5	0.0178	0.0010	0.4000

Table 2. Parameters of dynamic load model

R_2	X_2	X_1	α	P	$s(0)$	T'_{dot}	T_{jL}	K_{pq}
0.0186	0.2524	0.2505	1.6696	3.774	0.0500	16.00	0.0022	0.2326

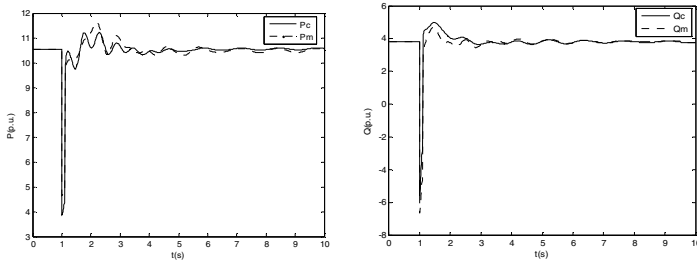


Fig. 5. Fitting effect of active and reactive power curves on fictitious bus

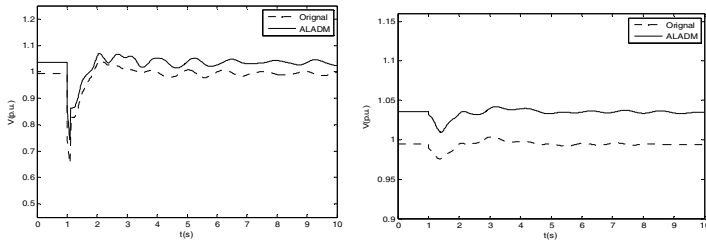


Fig. 6. Voltage of Bus30 under two tests

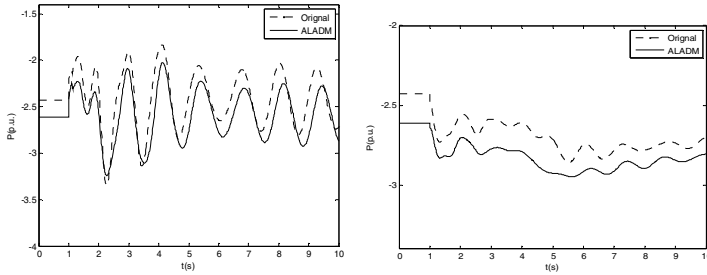


Fig. 7. Active injection power of Bus30 under two tests

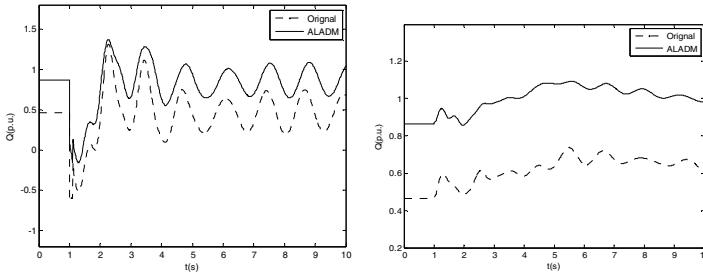


Fig. 8. Reactive injection power of Bus30 under two tests

Errors of voltage between model system and original system are calculated. The maximum errors of voltage amplitude and phase angle are less than 0.06p.u. and 0.1rad at Bus30.

Fig.5 gives the fitting effect of the dynamic load parameters on fictitious bus obtained by PDGA-RLS method. The above results show that the proposed ALADM can effectively simulate the original system.

To evaluate the equivalence of dynamic performance between model system with ALADM and original system, states on boundary buses, i.e. P, Q and V are calculated under the following two different typical operations. As limited space, only results of Bus30 which have the largest error are given. The voltage, active and reactive injection power are shown in Fig.6 to Fig.8.

Test 1: Three-phase short circuit fault lasting for 0.1s occurs on the line between bus 9 and bus 23.

Test 2: The output of unit 2 whose capacity is the largest in this power system, decrease 50% from 1st second. The results are shown in Fig. 8.and Fig.9.

It can be seen that although the load area has been simplified greatly, the external system and boundary line still keep their original characteristics well. This simulation of load modeling and network reduction based on the complex EPRI 36-bus power system shows that the methodology proposed in this paper has great potential in practice.

4 Conclusions

This paper presents a new methodology of using learning techniques based on WAMS to construct power system load models alongside network reduction. The ALADM, which is suitable for any real power system, is proposed and a PDGA-RLS algorithm is developed to optimize the parameters of this model. Simulation results show that this methodology offers a powerful modeling approach which can find a highly precise model to represent the load area in real power systems. As the data requested by the learning algorithm for constructing the equivalent model can be measured easily in the real power systems, it should be a practical technique with a great potential for exploitation in power system load modeling.

Acknowledgments. This work was supported in part by National HI-Tech Research and Development Program of China No.2006AA03Z209, Key Project of Chinese Ministry of Education No.107128, Program for NCET No.06-0643, and National Basic Research Program of China No.2004CB217906.

References

1. IEEE Task Force on Load Representation for Dynamic Performance: Bibliography on Load Models for Power Flow and Dynamic Simulation. IEEE Trans. Power Syst. 10, 523–538 (1995)
2. Yuan, Y., Ju, P., Li, Q., Wang, Y.Z.: A Real-time Monitoring Method for Power System Steady State Angle Stability Based on WAMS. In: Power Engineering Conference, pp. 761–764. IEEE Press, New York (2005)

3. Zhang, J., Yang, H.M., Li, K.: Power System Aggregate Load Area Model Based on Wide Area Measurement System. *Automation of Electric Power System* 31, 17–21 (2007)
4. Shen, S.D.: *Power System Identification*. Tsinghua University Press, Beijing (1993)
5. Ma, J.T., Wu, Q.H.: Generator Parameter Identification using Evolutionary Programming. *Int. J. Electr. Power Energy Syst.* 17, 417–423 (1995)
6. Goldberg, D.E.: *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Boston (1989)
7. Wu, Q.H., Ma, J.T.: Genetic Search for Optimal Reactive Power Dispatch of Power Systems. In: 1994 Int. Conf. Control, pp. 717–722. IEEE Press, New York (1994)
8. Power System Optimal Reactive Power Dispatch using Evolutionary Programming. *IEEE Trans. Power Syst.* 10, 1243–1249 (1995)
9. Linkens, D.A., Nyongesa, H.O.: Genetic Algorithms for Fuzzy Control. *IEE Proceedings-Control Theory and Applicat.* 142, 161–176 (1995)
10. Yin, X., Germang, N.: Investigations on Solving the Load Flow Problem by Genetic Algorithms. *Electrical Power Syst. Res.* 22, 151–163 (1991)
11. Ju, P., Handschin, E., Karlsson, D.: Nonlinear Dynamic Load Modeling: Model and Parameter Estimation. *IEEE Trans. Power Syst.* 11, 1689–1697 (1996)
12. IEEE Task Force on Load Representation for Dynamic Performance: Load Representation for Dynamic Performance Analysis. *IEEE Trans. Power Syst.* 8, 472–482 (1993)
13. IEEE Task Force on Load Representation for Dynamic Performance: Standard Load Models for Power Flow and Dynamic Performance Simulation. *IEEE Trans. Power Syst.* 10, 1302–1313 (1995)
14. Chang, A., Adibi, M.M.: Power System Dynamic Equivalents. *IEEE Trans. Power App. Syst.* PAS-89, 1737–1744 (1970)
15. Wang, L., Klein, M., Yirga, S., Kundur, P.: Dynamic Reduction of Large Power Systems for Stability Studies. *IEEE Trans. Power Syst.* 12, 889–895 (1997)
16. Cao, Y.J., Wu, Q.H.: A Cellular Automata Based Genetic Algorithm and its Application in Mechanical Design Optimization. In: 1998 Int. Conf. On Control, pp. 1593–1598. IEE, London (1998)
17. CEPRI, PSASP6.24 User Manual (2003)

Optimal Preventive Maintenance Inspection Period on Reliability Improvement with Bayesian Network and Hazard Function in Gantry Crane

Gyeondong Baek¹, Kangkil Kim², and Sungshin Kim¹

¹ School of Electrical Engineering, Pusan National University, 627706 Busan, Korea
{gdbaek, sskim}@pusan.ac.kr

² Equipment & Maintenance, Hajin Transportation Co., LTD, Pohang, Korea
kangkkim@hanjin.co.kr

Abstract. In this study, using Bayesian network about degradation model applied practically in gantry crane, the optimal preventive maintenance inspection period is suggested to improve the reliability of the parts. Central to this paper are two ideas. Bayesian network serves to indicate causal relation of the degradation units, and degradation of each unit is defined hazard function. Experimental results are presented to prove that the increase in degradation rate is due to the relation of parts. Proposed analysis method provides a stepping stone for developing basic technique for designing scheduled maintenance under uncertainly failure information.

Keywords: Bayesian network, Hazard function, Failure rate, Maintenance, Degradation model, Gantry crane.

1 Introduction

The gantry cranes occupy a crucial role within container terminal. A container terminal is a facility where containers are transshipped between different ships or yard tractors. And the gantry cranes, like the one shown in Fig. 1, are types of crane which lift objects by a hoist which is fitted in a trolley and move out along the rails to place the containers on the ship. As of present, marine transport accounted for 75 percent of total trade volume in the world. So the concern with automated container terminal design technology has been growing. Automated container terminal is an efficient operating system supported by cutting-edge technology. But in repair policy perspective, complex system has much cost to verify and test the unit [1-2]. For this reason, improving the maintenance effectiveness of crane can be extremely valuable.

Maintenance has always been important to the industry as it affects the performance. In general, maintenance is divided in corrective and predictive maintenance in ISO/SS 13306 standardization, predictive maintenance is further divided into time-based maintenance and condition-based maintenance in Fig. 2. It is a necessary action to sustain the performance, reliability and safety of the unit. To describe process of degradation units, it has been proposed hazard function. The bathtub-shaped hazard

function is described in nearly standard reliability [3-6]. And to represent causal relationships of units, it has been proposed Bayesian network. Bayesian network can provide probabilistic method to deal with uncertainty. A. Bobbio provides a helpful summary of the analysis of dependable systems by mapping fault trees into Bayesian network [7].

The main objective of this paper is to investigate the probabilistic relationships among units of gantry crane. With regard to dependable system, preventive maintenance inspection period relied on causal relation of the degradation units. The following section provides a structure of gantry crane, and faults of unit. Section 3 presents fault mechanism based on experience, and Bayesian network for extracting knowledge, and explains how hazard function is combined into each unit. Section 4 shows experimental result, and contains concluding remarks. Experimental data are used to compare with degradation time.



Fig. 1. A gantry crane in the Hanjin Transportation Co., LTD

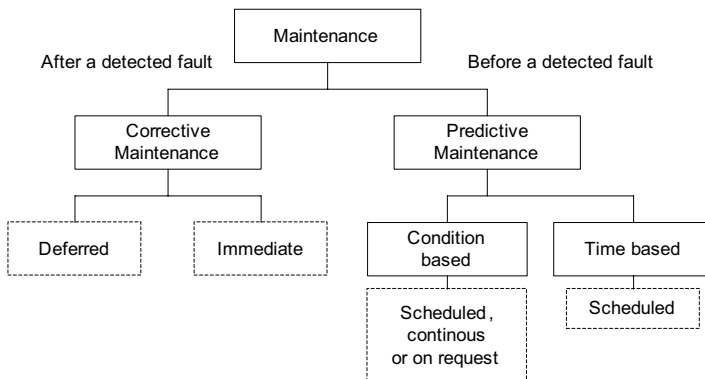


Fig. 2. ISO/SS 13306 standardization of maintenance terminology

2 Structure of Gantry Crane System and Fault of Unit

A gantry crane has many units. Their relations are portrayed in considerable detail, as shown in Fig. 3a. It is difficult that potential faults in gantry crane can be categorized. To find relation of faults easily, this paper will be limited to consideration of units, as shown in Fig. 3b. Trolley means a unit that travels on the bridge rails and carries the hoisting mechanism. Hoist means a system of power-driven drums, gears, cables, chains, or hydraulic cylinders capable of lifting and lowering a load. Spreader denotes a lifting device used to distribute forces. Twist lock is a function that locks between spreader and container using con-unit.

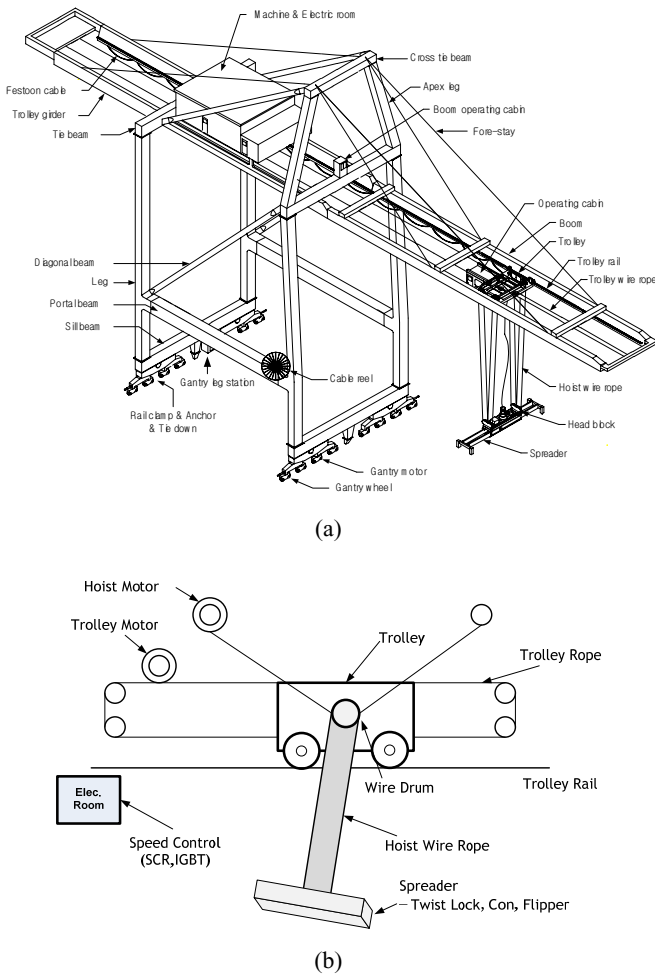


Fig. 3. Structure of a gantry crane: (a) units of general gantry crane system, (b) units related on hosting operation

The principal maintenance is concerned with hoist wire rope, and hoist motor. Maintenance inspection period of each unit is 1,800 hours, and 25,000 hours based on manufacturer. Since there are internal disturbances in the field, the causal relation caused a failure within maintenance inspection period. A failure within period is shown in Table 1. So it stands to reason that there are faults within maintenance inspection period.

Fault cases in Fig. 4 were divided into two groups: corrective maintenance and preventive maintenance. Causal relation in Fig. 5 is concerned with units of preventive maintenance. Among them are the following: hoist wire rope is influenced by 4 units, and hoist motor is influenced by 4 units.

Units of Preventive Maintenance

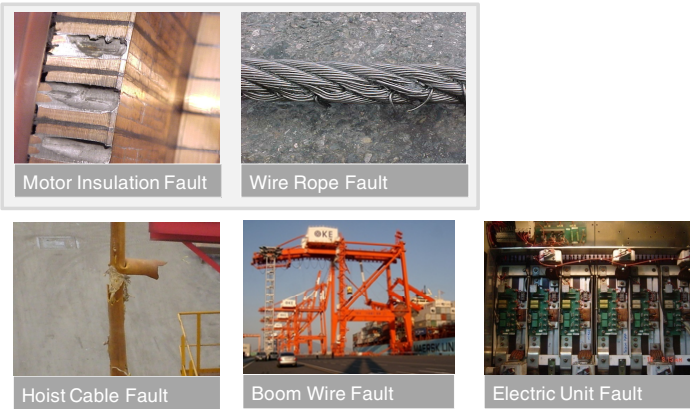


Fig. 4. Two fault cases: corrective maintenance and preventive maintenance

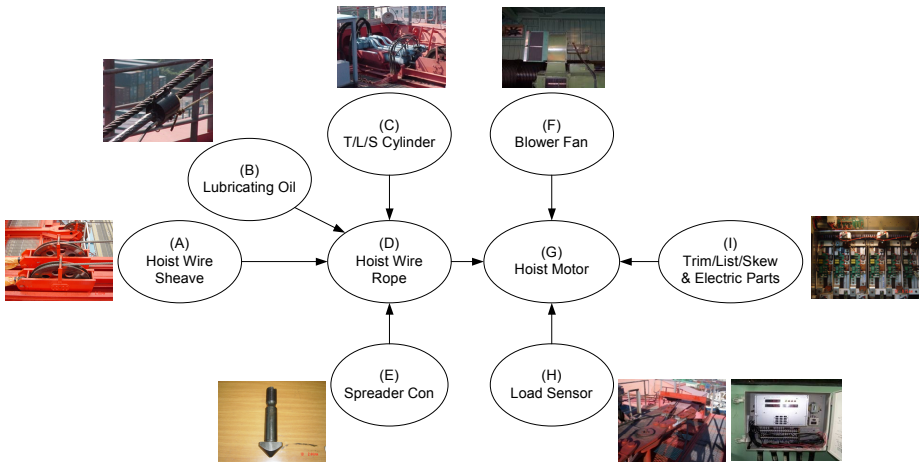


Fig. 5. Causal Modeling of faults in gantry crane

Table 1. Replacement day and used time of hoist wire rope on gantry crane

Types	Replacement							
	1	2	3	4	5	6	7	8
G/C101	98.8.8	99.3.8	99.11.8	00.12.8	01.11.23	02.11.25	03.10.1	04.7.2 6
Used (h)	1700	1145	1316	1865	1794	1653	1825	1781
G/C103	98.8.15	99.2.20	99.10.22	00.10.20	01.11.3	02.8.26	03.5.13	04.1.1 8
Used (h)	2000	1088	1376	1822	1867	1878	1891	1592

3 Combined Bayesian Network and Hazard Function

A Bayesian network is popular representation for uncertain expert knowledge in expert systems [8-9]. Bayesian networks learn about causal relationships. Knowledge of causal relationships allows us to make predictions. This section is devoted to the detailed method of degradation model for optimal preventive maintenance inspection period as shown in Fig. 6. The maintenance Bayesian network consists of two parts: prior probability and likelihood probability. Define the prior probability of degradation by

$$F(t) = 1 - e^{-H(t)} \tag{1}$$

Where $F(t)$ is the cumulative distribution function of time to failure, and $H(t)$ is the cumulative hazard function. This probability is defined according to:

$$F(t) = \lambda \left(1 - e^{-\left(\frac{t}{\eta_1}\right)^{\beta_1}} \right) + (1 - \lambda) \left(1 - e^{-\left(\frac{t}{\eta_2}\right)^{\beta_2}} \right), \quad t > 0 \tag{2}$$

Where β_1, β_2 are weibull shape parameter, η_1, η_2 is scale parameter, and λ is mixing parameter. Scale parameter is used to time of early failure and wear-out failure. For example, T/L/S electric unit has 48 hours in early failure, and 300 hours in wear-out failure as shown in Fig. 7. The parameters of hazard function are defined in Table 2.

Table 2. Hazard function parameters to describe self-degradation. (the effect of B neglecting).

	β_1	β_2	$\eta_1 (h)$	$\eta_2 (h)$	λ
A	5	5	48	3000	0.01
C	5	5	48	600	0.05
D	5	5	48	1800	0.01
E	5	5	48	35000	0.001
F	5	5	48	3000	0.01
G	5	5	48	25000	0.001
H	5	5	48	1000	0.01
I	5	5	48	300	0.05

To calculate $F_{bayes}(t)$, Table 3.a and Table 3.b show the likelihood tables of the relevant evidence based on operator's experience. $F_{bayes}(D)$, probability of hoist wire rope, is determined according to:

$$F_{bayes-D}(D) = \sum_{A,C,D,E,F,G,H,I} F(A,C,D,E,F,G,H,I) \tag{3}$$

Then,

$$\begin{aligned} F_{estimated-D}(t) &= \Pr[F_D(t) \cup F_{bayes-D}(t)] \\ &= F_D(t) + F_{bayes-D}(t) - \Pr[F_D \cap F_{bayes-D}] \\ &= F_D(t) + F_{bayes-D}(t) - \Pr[F_D | F_{bayes-D}] F_{bayes-D}(t) \\ &= F_D(t) + (1 - \Pr[F_D | F_{bayes-D}]) F_{bayes-D}(t) \\ &= \rho \end{aligned} \tag{4}$$

Where $(1 - \Pr[F_D | F_{bayes-D}]) = 1 - e^{-\theta F_D(t)}$ mean that the effect of internal disturbance is smaller until wear-out failure. So estimated using time of is determined according to:

$$F^{-1}(\rho) = t_{estimated-D} \tag{5}$$

Therefore, degradation time T_D is as follows:

$$T_D = t_{estimated-D} - t_{used-D} \tag{6}$$

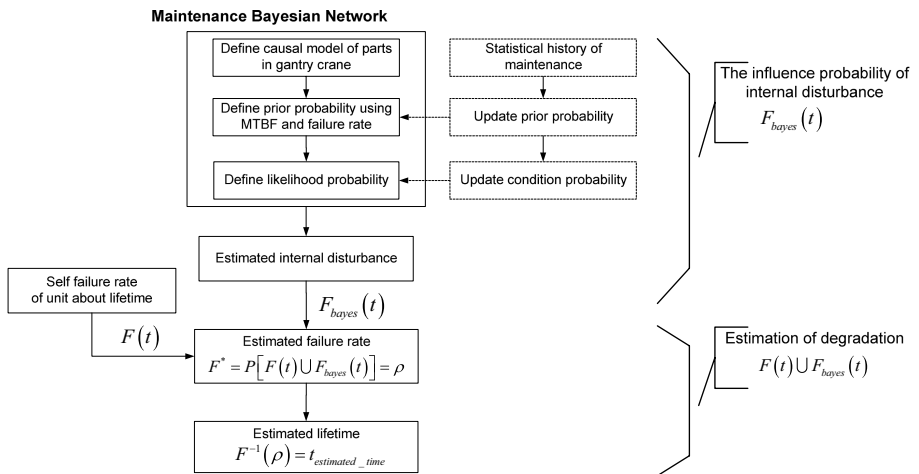


Fig. 6. Proposed degradation model to optimize preventive maintenance inspection period

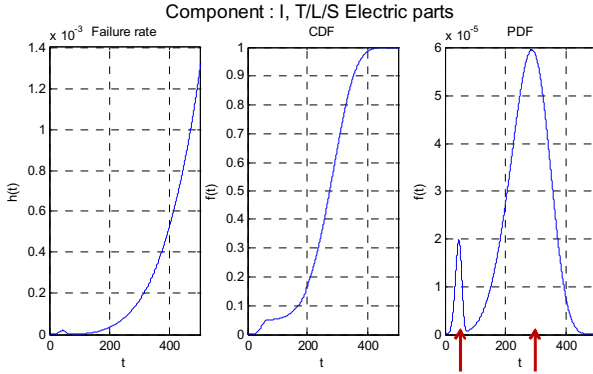


Fig. 7. T/L/S electric unit of gantry crane which has early failure and wear-out failure

Table 3. Effect of three units to hoist wire rope. (the effect of B neglecting).

Hoist wire sheave	T/L/S cylinder	Spreader con	Hoist wire rope
T	T	T	0.1
		F	0.6
	F	T	0.34
		F	0.7
F	T	T	0.6
		F	0.8
	F	T	0.7
		F	0.9

Table 4. Effect of four units to hoist motor

Hoist wire rope	Blower fan	Load sensor	T/L/S elec. unit	Hoist motor
T	T	T	T	0.1
			F	0.3
		F	T	0.6
			F	0.7
	F	T	T	0.6
			F	0.7
		F	T	0.7
			F	0.8
F	T	T	T	0.35
			F	0.65
		F	T	0.75
			F	0.65
	F	T	T	0.75
			F	0.75
		F	T	0.85
			F	0.9

4 Experimental Result and Conclusion

In our experiments, $F_{bayes-D}$ means internal disturbances of gantry crane. When internal disturbance is increased, estimated time is also increased as shown in Table 5. Bayesian network and bathtub-shaped hazard function are used to derive degradation time of equipment. Although available data is insufficient, analysis about system degradation is available by experience of operator and fault information from port. And proposed method will be available foundation technology to establish plan for preventive maintenance.

Table 5. Result of degradation time on hoist wire rope

Hoist wire rope Used time (h)	$F_{bayes-D}(t)$	Estimated time (h) (min, max)	Degradation (h) (min, max)
400	[0.1, 0.9]	(663, 1018)	(263, 618)
800	[0.1, 0.9]	(909, 1257)	(109, 457)
1200	[0.1, 0.9]	(1300, 1789)	(100, 589)

References

1. Lin, Y.J., Hung, C.Y., Hung, T.C.: Voltage Rise Due to Regenerative Braking of DC Machines Associated with Gantry Cranes at Kaohsiung Harbour. In: IEEE Canada Electrical Power Conference, pp. 452–455. IEEE Press, New York (2007)
2. Solihin, M.I., Wahyudi, Albaqul, A.: Development of Soft Sensor for Sensorless Automatic Gantry Crane Using RBF Neural Networks. In: IEEE Conference on Cybernetics and Intelligent Systems, pp. 1–6. IEEE Press, New York (2006)
3. Klutke, G.A., Kiessler, P.C., Wortman, M.A.: A Critical Look at the Bathtub Curve. IEEE Trans. on Reliability 52, 125–129 (2003)
4. Hjarth, U.: A Reliability Distribution with Increasing, Decreasing, Constant and Bathtub-Shaped Failure Rates. Technometrics 22, 99–107 (1980)
5. Dimitrakopoulou, T., Adamidis, K., Loukas, S.: A Lifetime Distribution with an Upside-Down Bathtub-Shaped Hazard Function. IEEE Trans. on Reliability 56, 308–311 (2007)
6. Xie, M., Lai, C.D.: Reliability analysis Using an Additive Weibull Model with Bathtub-shaped Failure Rate. Reliability Engineering and System Safety 52, 87–93 (1995)
7. Bobbio, A., Portinale, L., Minichino, M., Ciancamerla, E.: Improving the Analysis of Dependable Systems by Mapping Fault Trees into Bayesian Networks. Reliability Engineering and System Safety 71, 249–260 (2001)
8. Chickering, D., Geiger, D., Heckerman, D.: Learning Bayesian Networks: Search Methods and Experimental Results. In: Fifth Conference on Artificial Intelligence and Statistics, pp. 112–128. IEEE Press, New York (1995)
9. Sheu, S.H., Yeh, R.H., Lin, Y.B., Juang, M.G.: A Bayesian Approach to an Adaptive Preventive Maintenance Model. Reliability Engineering and System Safety 71, 33–44 (2001)

Application of RBF Network Based on Immune Algorithm to Predicting of Wastewater Treatment

Hongtao Ye^{1,2}, Fei Luo¹, and Yuge Xu¹

¹School of Automation Science and Engineering, South China University of Technology, 510641 Guangzhou, China

²Department of Electronic Information and Control Engineering, Guangxi University of Technology, 545006 Liuzhou, China
yehongtao@126.com

Abstract. Wastewater treatment is a nonlinear, time-varying and time-delay process. It is difficult to establish exact mathematic model. A novel radial basis function (RBF) neural network model based on immune algorithm (IA) is presented in this paper. It combines the merits of IA and neural network. The IA is used to determine the hidden layer clustering centers of RBF neural network. The wastewater treatment process is established with the novel RBF neural network. Simulation results prove that the method has the advantages of less computation and higher precision.

1 Introduction

Wastewater treatment is a complex and nonlinear biological reaction process. Exact mathematic model includes many biochemical reactions. Because it is difficult to obtain reliable kinetic parameters, the predicting of wastewater treatment is not exact.

The emergence of intelligent algorithms makes it possible to establish complex system models [1,2]. Andrea [3] applied BP neural network to establish model of wastewater treatment process. Yu [4] adopted adaptive fuzzy neural network to establish activated sludge treatment process model, achieving satisfactory results. Wan [5] combined artificial neural network and genetic algorithm (GA) to predicting of effluent water quality.

However, due to BP neural network use of steepest descent method to search for the optimal solution, it can not guarantee that the error function converge to the global optimum. Adaptive fuzzy neural network inevitably exists structure identifying problems. Neural network has some shortcomings such as the longer training time and the random choice structure of network. GA can not guarantee that the training of the network will not be local minimum region.

The IA is inspired by the human immune system which is a remarkable natural defense mechanism that learns about foreign substances [6]. It is one of the evolutionary algorithms that imitate the immune system to solve the optimization problem. Although IA is very similar to GA, there are essential differences owing to the memory education system and production system for various antibodies. IA is widely applied in pattern recognition and data clustering [7].

This paper combines the merits of IA and RBF neural network. The IA is used to determine the hidden layer clustering centers of RBF neural network. The novel RBF neural network model is used to predicting of effluent water quality.

2 Wastewater Treatment Process

A general overview of wastewater treatment process is shown in Fig.1 .Sewage is collected by local pumping stations to the plant. Influent firstly passes through an anaerobic tank and anoxic tank in which nitrate is reduced to nitrogen and organic carbon is consumed. In the aeration tank, further consumption of organic carbon occurs, and ammonia is converted to nitrate. The secondary tank returns a concentrated stream of activated sludge to the anaerobic tank while passing clarified effluent through the filter to the disinfection tank for inactivation of harmful microorganisms.

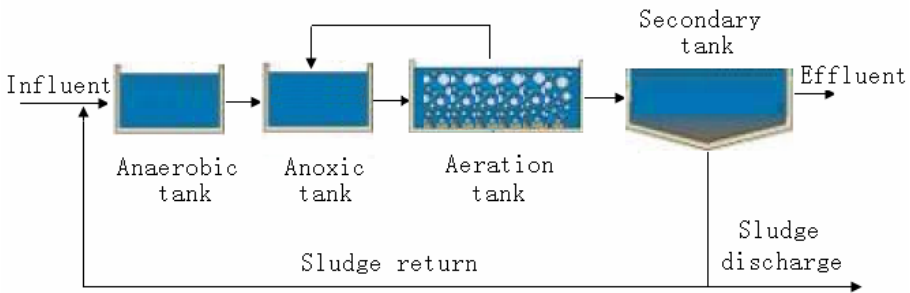


Fig. 1. Schematic diagram of wastewater treatment process

3 Algorithm Analysis

3.1 RBF Neural Network

RBF neural network structure includes input layer, hidden layer and output layer [8]. The input

$$X = [x_1, x_2, \dots, x_N] \tag{1}$$

$$x_i = [x_{i1}, x_{i2}, x_{iP}]^T \tag{2}$$

The output

$$Y = [y_1, y_2, \dots, y_O] \tag{3}$$

$$y_i = W_i^T G = \sum_{j=1}^m w_{ij} g_i, i = 1, 2, \dots, O \tag{4}$$

$$W_i = [w_{i1}, w_{i2}, \dots, w_{im}]^T, i = 1, 2, \dots, O \tag{5}$$

$$G = [g_1, g_2, \dots, g_m]^T \tag{6}$$

Where G is radial basis function.

$$g_j = \psi_j \frac{\|x - c_j\|}{\sigma_j} = e^{-\frac{\|x - c_j\|^2}{\sigma_j^2}}, j = 1, 2, \dots, m \tag{7}$$

Where c_j is the center of basis function, and σ_j is the width of radial basis function around center.

3.2 RBF Network Based on IA

Definition 1: Affinity

The affinity represents the matching level between the antigen Ag and the antibody Ab . The affinity can be computed through

$$a = \frac{1}{1 + \|Ab - Ag\|} \tag{8}$$

Definition 2: Similarity

The similarity represents the matching level between the antibody Ab_i and the antibody Ab_j . The similarity can be computed through

$$s = \frac{1}{1 + \|Ab_i - Ab_j\|} \tag{9}$$

The IA is used to determine the hidden layer clustering centers of RBF neural network. The steps of RBF network based on IA can be described as follows:

Step1: Initialize hidden layer centers. N random data are generated as hidden layer centers cluster C . Antibodies correspond to C .

Step2: Antigens correspond to x_i . Each input x_i is operated as follows:

Step2.1: Calculate the affinity values of C with x_i . A better antibody has a higher affinity with the antigen.

Step2.2: Select the highest affinity antibody, and reproduce N_1 copies as cluster K . Then mutate K as cluster K_1 .

Step2.3: Calculate the affinity values of K_1 with x_i . Select N_2 higher affinity antibodies as cluster K_2 , then eliminate the antibodies which affinity is lower than threshold as new cluster K_3 .

Step2.4: Clone restrain. In order to boost up the multiplicity, eliminate the antibodies which similarity is larger than threshold as new cluster K_4 .

Step2.5: Combine K_4 which is produced by x_i as new cluster C_1 .

Step3: Clone restrain C_1 . Eliminate the antibodies which similarity is larger than threshold as new cluster C_2 .

Step4: Stop and go to Step5 if the stopping criterion is satisfied, otherwise add random antibodies to C_2 and return to Step2;

Step5: C_2 are the hidden layer centers, then establish RBF network.

4 Simulation and Analysis

Effluent water quality is the most important criterion of wastewater treatment effect. Effluent water quality is affected by a number of factors. Effluent NH_3 is chosen as the output of RBF network. Through in-depth researching the removal mechanism of nitrogen and phosphorus of wastewater, dissolution oxygen (DO), PH, chemical oxygen demand (COD), total phosphorus (TP) and influent NH_3 are chosen as the input of RBF network.

The simulation was implemented with MATLAB version 6.5 and its nnet toolbox. Simulation data came from the Guangzhou Lijiao sewage treatment plant. As wastewater biological reaction process is slow, the data will not vary much in a short time. The data sampling period is 10 minutes. The simulation results presented in Fig. 2 and Fig. 3 illustrate that the RBF network based on IA performs quite well for the prediction of effluent NH_3 . The prediction precision of RBF network based on IA is higher than RBF network. Fig. 4 and Fig. 5 show the convergence of RBF and RBF based on IA training. The training process is activated to a performance target of 0.005. RBF network training epoch is 65, while RBF network based on IA is 33. The RBF network based on IA has a faster speed of training.

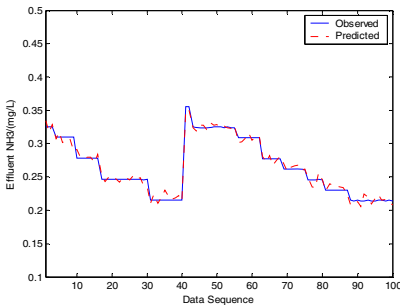


Fig. 2. RBF network prediction for NH_3

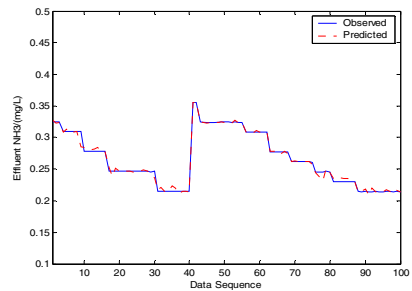


Fig. 3. RBF based on IA prediction for NH_3

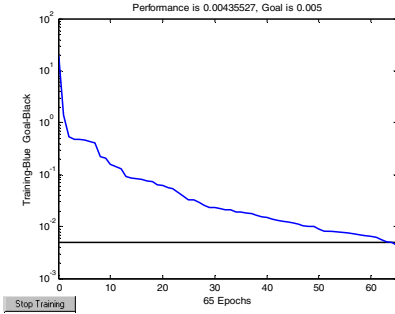


Fig. 4. Convergence of RBF training

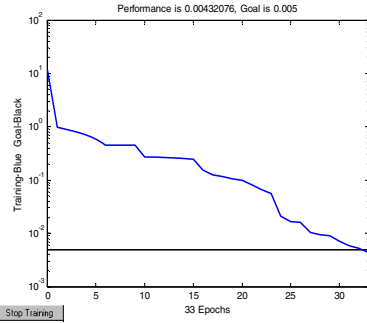


Fig. 5. Convergence of RBF based on IA training

5 Conclusions

Traditional algorithms to predicting of the wastewater treatment process have many defects. The simulation illustrates that RBF neural network based on IA is effective. The method has the advantages of less computation and higher precision. Certainly, choosing different parameters will affect the result of this algorithm, and how to realize adaptive selection of the parameters is a question which worth further study.

Acknowledgments. This research was supported by National NSF of China #60774032, Special Research Fund of Ministry of Education of China for College Doctoral Subjects (Project for Young Scholars) #20070561006 and Guangxi University of Technology master fund project #500537.

References

1. Hanbay, D., Turkoglu, I., Demir, Y.: Prediction of Wastewater Treatment Plant Performance Based on Wavelet Packet Decomposition and Neural Networks. *Expert Systems with Applications* 34, 1038–1043 (2008)
2. Hamoda, M.F., Ghusain, I.A., Hassan, A.H.: Integrated Wastewater Treatment Plant Performance Evaluation Using Artificial Neural Networks. *Water Science Technology* 40, 55–65 (1999)
3. Andrea, G.C., Harold, V.J., Vladimir, N.: Sludge Bulking Analysis and Forecasting: Application of System Identification and Artificial Neural Computing Technologies. *Water Research* 25, 1217–1224 (1991)
4. Yu, Y., Qiao, J.F.: Modeling Study of Activated Sludge Process Based on ANFIS. *Computer Engineering* 32, 266–268 (2006)
5. Wan, T.J.: An Application of Artificial Neuromolecular System for Effluent Quality Prediction of Wastewater Treatment Plant. *Chinese Instrument Environment Engineering* 10, 155–162 (2000)

6. Jiao, L.C., Du, H.F., Liu, F.: Immune Optimization: Computing, Learning and Recognition. Science Press, Beijing (2006)
7. Deng, J.Y., Mao, Z.Y., Luo, Y.H.: Pattern Recognizing Algorithm Based on Artificial Immune Network. Journal of South China University of Technology: Natural Science Edition 36, 99–103 (2008)
8. Sun, Z.Q., Zhang, Z.X., Den, Z.D.: Intelligent Control Theory and Technology. Tsinghua University Press, Beijing (1997)

HLA-Based Emergency Response Plan Simulation and Practice over Internet

Wan Hu¹, Hong Liu², and Qing Yang²

¹ School of Mechatronics Engineering, University of Electronic Science & Technology of China, 610054 Chengdu, China

² Institute of Systems Engineering, Huazhong University of Science and Technology, 430074 Wuhan, China
wanhu@uestc.edu.cn

Abstract. Efficient emergency response for disasters needs systematic response preparedness and plan. Distributed computer simulation and practice drilling can help to perfect the emergency response plan and train the participants. In this paper, a distributed simulation and practice framework which realizes extending HLA/RTI to Internet based on Grid service is proposed. The framework aims to the advantage of Grid technology as well as the reusability and interoperability of simulation modules. The results of experiments of the prototype indicate the feasibility of the framework, which provide a platform for emergency response drilling distributed simulation over Internet. At the end of paper, the future development plan has been discussed.

Keywords: Emergency response, Simulation and practice, Grid service, HLA, WS-Notification.

1 Introduction

Disasters such as floods, earthquakes, and outbreak of epidemics pose a greater risk to populations [1,2]. Efficient emergency response for both natural and man-made disasters can reduce the damage and casualties. Thus, it is important for emergency planners to take a broad approach to disaster preparedness and plan for the consequences from disasters [4,5]. However, the emergency response plan is typically composed by the idea of diverse stakeholders and subject matter experts and it should be evaluated and revised through practical drilling repeatedly to be optimized. Computer modeling and simulation can help to train response participation people and ensure that the planning and evaluation process is systematic, logical, and complete in a low cost, risk-free setting.

Disaster preparedness planning operates in a convoluted, confused, and fragmented environment. It involves a variety of governmental and no-governmental agencies, including decision-making body, fire, emergency medical services, hospitals, police, and public health, with overlapping jurisdictions and competing agendas and interests [3,4,5]. These agencies, which involved in simulations of Disasters response, are commonly geographically distributed and subjected to different organizations. Since

their private computer networks need not exchange large data, special net lines are seldom constructed. Therefore the public Internet is the appropriate way if they are managed to interconnect to set up a distributed computer simulation.

Currently in most research projects of disaster emergency response simulation & practice, for convenient intercommunication between simulation model components of diverse agencies, they are centralized in a simulation center. However, since simulation drilling of disaster emergency response often need to access local data of the agencies and refers human-computer interaction, it's impracticable to run a simulation in a simulation center and inconvenient to assemble all related staff of multi-agencies every time when to organize a simulation drilling.

The High Level Architecture (HLA) is an IEEE standard for simulation and modeling and provides application developers with a powerful framework for distributed simulation reuse and interoperability. However its design was not intended to support software applications that need to integrate instruments, displays, computational and information resources managed by diverse organizations [6]. In order to run a distributed simulation over the Wide-Area-Network (WAN) using the IEEE HLA/RTI directly, special arrangements have to be made beforehand to ensure the availabilities of the required hardware and software. Such arrangements are typically made with a centralized control or simply within an organization, because inter-organizational sharing of resources involves issues such as security [7].

The advent of computing Grid technology enables the use of distributed computing resources and facilitates the secure access of geographically distributed data. It provides an unrivalled opportunity for facilitating the large-scale distributed simulation.

The Open Grid Services Architecture (OGSA), developed by The Global Grid Forum, aims to define a common, standard, and open architecture for grid-based applications. Web services provide an approach to distributed computing with application resources provided over networks using standard technologies. It is based on a defined set of technologies, supported by open industry standards, that work together to facilitate interoperability among heterogeneous systems. Web Services Resource Framework (WSRF), a specification developed by OASIS, extends Web services to stateful services, which OGSA requires.

WS-Notification is a family of related specifications (including WS-BaseNotification, WS-BrokeredNotification, and WS-Topics) that define a standard Web services approach to notification using a topic-based publish/subscribe pattern and it had been approved to be OASIS standard in 2006. WS-BaseNotification defines the Web services interfaces for NotificationProducers and NotificationConsumers. WS-Topics define a mechanism to organize and categorize items of interest for subscription known as "topics." WS-BrokeredNotification defines the Web services interface for the NotificationBroker. The Globus Toolkit 4.0 (GT4) is a software toolkit developed by The Globus Alliance. It is an implementation of OGSA and a sort of de facto standard for the Grid community. GT4 currently implements part of WS-Notification including effective topic-based notification [8].

In recent years, some remarkable research work focus on combining Grid Technology and HLA for simulations to take advantages of both. Katarzyna Zajac, etc. gave the idea of a three-level approach to building the Grid services for HLA-based applications. They also care about supporting execution of HLA distributed interactive simulations in a Grid environment and federate migration [6,9].

Stephen J. Turner etc. propose a distributed simulation framework, called HLA Grid. The framework uses a Federate-Proxy-RTI architecture, which allows resources on the Grid to be utilized on demand by using Grid services. In HLA Grid, RTI services are exposed as Grid services and federates' RTI Ambassador call is translated to remote Grid service invocations. Correspondingly, federates' Federate Ambassador callback is also exposed as Grid services to be invoked by the RTI side [8]. Grid service invocation communicates via Simple Object Access Protocol (SOAP), which commonly bases on HTTP. So translating the Federates-RTI communication to Grid services invocations provides the feasibility that Federates-RTI communicates across the public WAN environment.

In this paper, in order to execute distributed simulations of emergency response plan over the Internet, we propose a distributed simulation framework which realizes extending HLA/RTI to Internet based on Grid services. In the proposed framework, it's maintained that RTI Ambassador call is encapsulated within Grid service, while a multithreading Grid service is designed. Federates' Federate Ambassador callback is implemented with WS-Notification. Thus simulation federates need not to be exposed as Grid service that lighten the development and running complexity of the system. Since the distributed applications run on Internet environment, the security problem is also considered in the framework. The implemented system provides a flexible and extensible environment, which can be used to set up the distributed simulation drilling of the emergency response planning over Internet.

The rest of this paper is organized as follows. In Section 2, the overall architecture of the proposed simulation framework is described. In Section 3, a distributed simulation environment is constructed using the design of Section 2. In Section 4, based on the environment, simulation example applications are executed to verify the environment and system, while conclusions and future works reside in Section 5.

2 Distributed Simulation System Design

2.1 Framework Overview

The basic components of simulation framework are showed in figure 1. The system is composed of two essential parts: RTI service side and simulation federates side.

In RTI service side, RTI execution has a corresponding Grid service. The Grid service may be in the same LAN environment with the RTI and communicate with RTI via common LRC (Local RTI Component). Grid service also provides services to be invoked by the remote simulation federate. The Grid service keeps multiple threads that correspond to simulation federate one by one. Each thread contains an instance of LRC, through which to interact with RTI.

The practical simulation application may involve series of simulation federates. Each simulation federate contains simulation code and a fake LRC component. The fake LRC component provides standard IEEE HLA interface to above simulation code. It invokes remote Grid service and accepts the notifications from the grid service to interact with RTI.

Communication authentication and encryption communication between Federate/Client side and RTI services side ensures secure conversation.

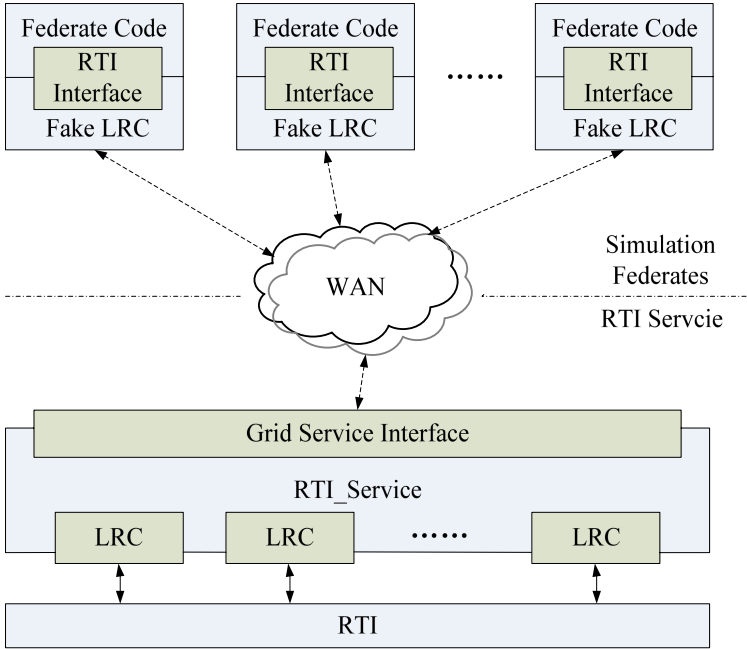


Fig. 1. Architecture of simulation based on Grid service

The system also contains an Index service. Index service provides registry service for Grid service and maintains information of RTI and the corresponding Grid service to be query by the fake LRC in federate side.

2.2 RTI-Federate Interaction Procedure

WS-Notification is used to realize Federate Ambassador callback, while RTI Ambassador call remains to be transferred to Grid services invocation. The Interaction procedure between federates and RTI is showed as Figure 2.

It's assumed that the Grid service invocations are on secure communication and both sides have been authenticated before.

When federate code calls RTIAmbassador interface function, the local fake LRC will encode the parameters to Grid service invocation format and then invoke remote Grid service. Grid service in RTI service side will decode the parameters back to RTIAmbassador call format and call the local common LRC, which corresponds with the federate, to realize interaction. The call return process is similar and converse. As showed in Figure 2, RTIAmbassador call procedure takes mode of synchronization, in order to not disturb the internal Time Management of RTI.

To realize FederateAmbassador callback, Grid service in initialization will firstly publish WS-topics and fake LRC in federate side will subscribe the WS-topics before joining the simulation federation. When LRC in Grid service receives FederateAmbassador callback, Grid service will encode the parameters to notification format and notify the corresponding federate. Whenever fake LRC in federate side receives the

notification, it decodes the parameters and calls the federate's corresponding FederateAmbassador function. In FederateAmbassador callback procedure, federates are running in asynchronous mode. When receiving notification, federates will be interrupted to handle it.

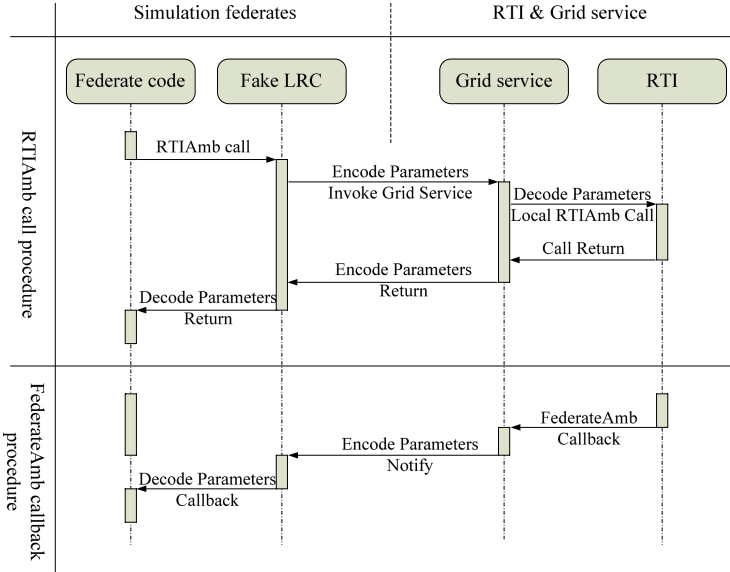


Fig. 2. Interaction procedure between federates and RTI

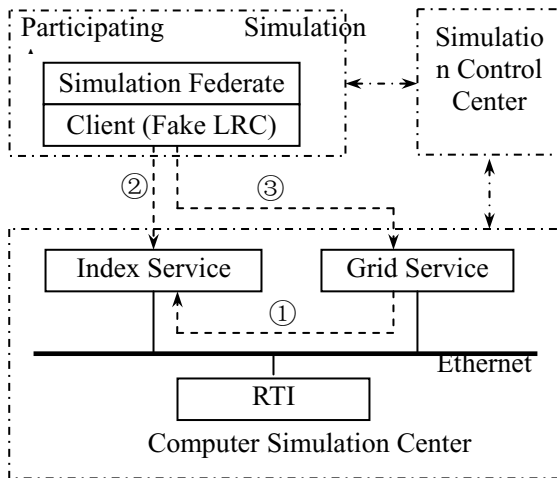


Fig. 3. Simulation application procedure

2.3 Practical Application Procedure

The practical procedure of running a emergency response simulation over Internet is showed in Figure 3. The application system contains three kinds of participators: the Computer Simulation Center (CSC), the Emergency Response Simulation Control Center (ERSCC) and the Participating Simulation Agency (PSA). Index service, Grid service and RTI can be deployed in CSC. ERSCC is responsible for the simulation coordination control and does not join the simulation. In each PSA, simulation model of itself runs locally. PSA communicates with CSC via Internet.

When CSC and PSAs receive the command of joining a simulation drilling practice, they will deploy and initialize the related components. ① During initialization, Grid service registers the information of corresponding RTI and itself in the Index Service, and publishes the WS-Topics. After initialization, CSC makes response to ERSCC. ERSCC then informs PSAs. ② During the initialization of Fake LRC component in PSA, it will inquire about appropriate RTI and corresponding Grid Service at Index Service before Grid Service invocation. ③ When Fake LRC firstly receives RTIAmbassador call from simulation code, it will invoke the corresponding Grid Service. Grid Service then begins a new threading, which keeps an instance of LRC. Once the federate joins the Federation, Fake LRC will subscribe the corresponding WS-Topics. Though the communication procedures showed as Figure 2, federates interact with RTI and enter simulation cycle. ④ At the end of simulation, federate calls the LRC in Grid Service to resign and destroy simulation federation. And the corresponding threading will be stopped.

3 Prototype Implementation

Based on above framework, a simplified simulation environment prototype for feasibility issues has been implemented. It's developed in Java. The Grid middleware runs the GT4, and DMSO RTI NG 1.3V6 is used.

At federates side, to realize the Fake LRC, the class `hla.rti1_3V6.RTIambassador` is rewritten and recompiled. In detail, the interface of the class methods is maintained, but the material realization content is rewritten. In constructor function, it mainly contains the Grid Service invocation initialization and WS-topics subscription. The other methods transfer RTIAmbassador call to Grid Service invocation. In addition, the class provides a callback for WS-Notifications. The callback interrupts the federate to handle corresponding FederateAmbassador callback.

A Grid Service named `RTI_Service` is developed based GT4. `RTI_Service` provides interface for service invocations in Fake LRC. The methods implementations mainly transfer Grid Service invocation to RTIAmbassador call. In order to handle FederateAmbassador callback, it takes charge of WS-Topics publish and notification. Designed as a multithread Architecture, it also manages the multiple instances of LRC.

Index Service is reduced. Fake LRC addresses `RTI_Service` directly.

4 Experiments and Results

In order to demonstrate the feasibility and evaluate system quality, two experimental tests are executed based on the prototype platform.

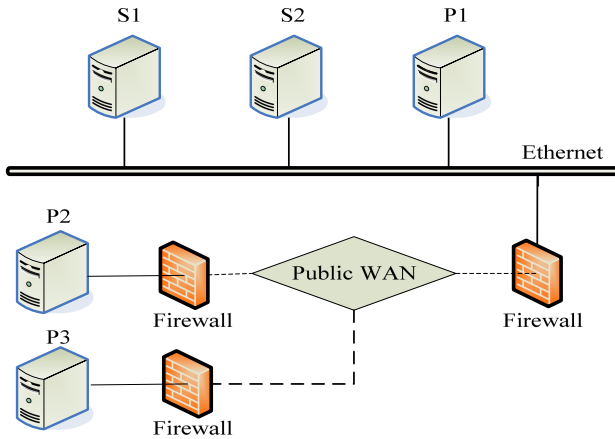


Fig. 4. Experiment hardware configuration

As showed in Fig4, the testing is done using the computational resource at Systems Engineering Institute (SEI) of Huazhong University of Science & Technology (HUST) in Wuhan China and other two workstations. Machines S1, S2 and P1 in SEI are inter-connected using Ethernet. The workstations P2, P3 connect with SEI via the public WAN of campus network of HUST and have an interval of 6 hops to SEI. Individual machines' specifications are shown in Table1.

Table 1. Specification of machines for experiments

	S1, S2	P1	P2, P3
Hardware	IBM(886MB256MBx2)	PentiumIV 2.0GHz256MB	PentiumIV 2.0GHz512MB
OS	Fedora™ Core2	Fedora™ Core2	Fedora™ Core2

In experiment 1, there are three components: RTI, RTI_Service and Federate1. RTI and RTI_Service are executed in SEI on machines S1, S2. Federate1 is executed on P2 and connects with RTI through Fake LRC-RTI_Service. Federate1 runs the Hello-Java demo program successfully.

Comparison experiments are done to analyze time-delay brought by the system structure. Federate1 is executed on P1 in the same LAN with RTI. The simulation time consumptions are recorded while federate interacts with RTI through either common LRC or Fake LRC-RTI_Service. In each step of the simulation loop of HelloJava program, there mainly are an updateAttributeValues, a timeAdanceRequest RTIAmbassador call and a timeAdvanceGrant FederateAmbassador callback. 50 steps of simulation loop are selected.

The experimental result data are showed in Table 2. The data of LAN network show that each step of simulation loop based on our prototype consumes more time of 420ms. That is mainly due to network transmission delay and encoding/decoding of

parameters cost. In Grid Service invocation, formed into XML and transported though SOAP, the communicated data is greatly increased. When it comes to campus public WAN, the time-delay is more remarkable because of network bandwidth limit.

Table 2. Data of time-consumption

Communication network	LAN		Public WAN
Interactive mode	LRC	Fake LRC-RTI_Service	Fake LRC-RTI_Service
Time-consumption (50steps)	610ms	21700ms	158000ms

In experiment 2, based on the prototype platform, a simplified emergency response scenario is simulated. Shown as Figure 5, the simulation federation contains three federates: Federate1 models the emergency decision command center actions, Federate2 models a medical institution, which include two departments: a hospital and medical transport service, and Federate3 models a drugs manufacturing enterprise. The emergency decision command center is responsible for connecting the potential productivity data and making the response plan.

The interactive data between federates is little. Firstly, the decision command center gets a call that several casualties of high severity at certain location is waiting for help. Through man-computer interaction in the center, it is determined how to service the casualty including dispatching an ambulance to the casualty using a specific selected route and then on to the hospital. Then the command information is sent to medical federate. When the hospital urgently needs some drugs, it requests the command center. The command center analyzes the demand information and sends the order to drugs manufacturing enterprise. When the hospital gets the drugs, the simulation drilling is over.

That experiment that simulating a simplified emergency response drilling further verifies the feasibility of the framework.

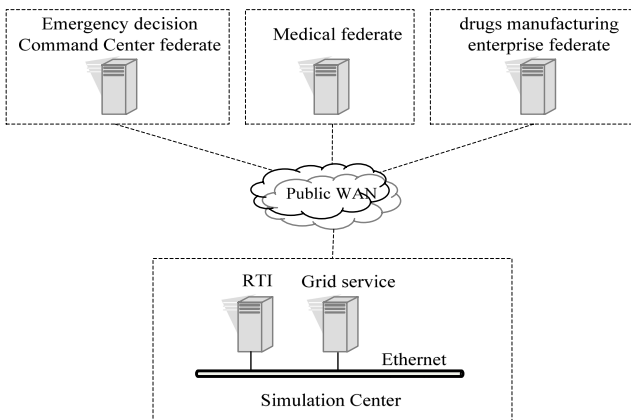


Fig. 5. A simple emergency response scenario simulation

5 Conclusion and Future Work

In this paper, we propose a distributed simulation framework, which realizes extending HLA to Internet environment based on Grid service. It provides a platform for disaster emergency response drilling distributed simulation over Internet. The framework takes the advantage of the security management of Grid to run distributed simulations that span several organizations over Internet and provides standard HLA interface for simulation federate components, considering reusability. WS-Notification is used to implement FederateAmbassador callback. Thus local federates need only get the Fake LRC component to connect with remote RTI, which provides flexibility and practicability to end-users.

The experiments of prototype indicate the feasibility of providing platform for the distributed simulation of the disaster emergency response drilling over Internet. However, since a Grid service is imported between the federate and RTI, additional time-consumption is incurred. That will be remarkable especially in Internet environment.

The presented work is a first step approach yet. Much work remains to be done to perfect the prototype system to fulfill the overall framework, such as adding security to Grid service and development of Index service.

Reducing the time-consumption of communication between Federate/RTI in the system need to be further researched to meet the real time requirement of some distributed simulation applications. The future work also includes fault-tolerance, considering robustness of the system.

Acknowledgments. Great thanks for the support from National Natural Science Foundation of China (NSFC, Grant 60773188) and China Postdoctoral Science Foundation (CPSF Grant 20080430961).

References

1. Pan American Health Organization: Natural Disasters: Protecting the Public's Health. Washington, DC (2000)
2. Levy, B.S., Sidel, V.W.: Terrorism and Public Health. Oxford University Press, Oxford (2003)
3. Making the nation safer: The role of technology and science in countering terrorism, <http://www.nap.edu/catalog/10415>
4. Emergency Responders: Drastically Underfunded, Dangerously Unprepared, http://www.cfr.org/pdf/Responders_TF.pdf
5. Nash, D.B.: Being Ready for an insidious Threat. *Managed Care* 12(1) (2003)
6. Zaja, K., Tirado-Ramos, C.A., Zhao, Z., Sloot, P.: Grid Services for HLA-based Distributed Simulation Frameworks. In: 1st European AcrossGrids Conference, pp. 147–154. Springer, Heidelberg (2003)
7. Xie, Y., Teo, Y.M., Cai, W., Turner, S.J.: Service Provisioning for HLA-based Distributed Simulation on the Grid. In: Workshop on Principles of Advanced and Distributed Simulation, pp. 282–291. IEEE Press, New York (2005)
8. The Globus Toolkit 4 Programmer's Tutorial (2006), <http://gdp.globus.org/gt4-tutorial/>
9. Zajac, K., Bubak, M., Malawski, M., Sloot, P.M.A.: Towards a Grid Management System for HLA-based Interactive Simulations. In: 7th IEEE International Symposium on Distributed Simulation and Real Time Applications, Delft, Netherlands, pp. 4–11 (2003)

Dynamic Cooperation Mechanism in Supply Chain for Perishable Agricultural Products under One -to- Multi

Lijuan Wang¹, Xichao Sun¹, and Feng Dang²

¹Department of Computer, Henan Agricultural University,
Zhengzhou 450002, China
wsk@henau.edu.cn

²Department of Information Management of Zhengzhou University,
Zhengzhou 450002, China

Abstract. This paper is to study dynamic cooperation mechanism in a two-stage perishable Agricultural products supply chain by applying grey system theory. The two-stage perishable agricultural products supply chain system is consist of a farmer and many companies. the grey game model is first set up by analyzing the characteristics of dynamic price. Then, the various factors of impacting a farmer and companies cooperation are analyzed. In doing so, the optimization equilibrium strategy and measure to maximize both sides profit under grey market price are discussed. The dynamic cooperation mechanism is designed to optimize the system. Finally, the reasonable and effective of dynamic cooperation mechanism are shown by an example.

Keywords: Dynamic cooperation mechanism, Perishable products, Grey matrix, Grey game, One -to-multi.

1 Introduction

Supply chain management has received considerable attention in the business-management literature ^[1]. Coordination has been a major research issue in the study of supply chain management. The cooperation mechanism or cooperation contract is the main means which coordinates and constraints every one of supply chain to operate to make the whole supply chain performance is optimal^[2-3]. In the supply chain for perishable agricultural products which is consist of Corporation & Farmer Mode, It is always to happen that the breaching of contract behavior in "Company & Farmer Household under present mechanism. This hinders seriously the development of Corporation & Farmer Mode. How to deal with their cooperation is the main key to improve the supply chain performance and to realize overall optimization. During recent years, the research of cooperation mechanism for Corporation & Farmer Mode under dynamic focuses on three central themes: (i) dynamic game^[4], (ii) contract, the theory of principal and agent, (iii) technology analysis of feed-back dynamic complexity combined with the theory of principal and agent^[5-13]. The above studies have provided tremendous help for SCM management. However, they did not research the cooperation issue from dynamic characteristics under strong uncertain marker price.

Due to the impact of natural disaster, market, national macro policies, and the future uncertain random fluctuation factors and so on, this made fresh agricultural products price very strong uncertain and the market price isn't judged quite accurately beforehand. This situation results in that the profits of "Corporation & Farmer", the perform-contract and the game results between "Corporation & Farmer" in actual perishable agricultural products supply chain don't judge quite accurately beforehand. According to the above problems existing, this paper applies the grey system theory to supply chain management (SCM) and focus on the aspect of price dynamic characteristics of perishable agricultural products.

The author has already made a series of studies in supply chain for perishable agricultural products: the first dynamic cooperation mechanism has been discussed in grey price under one-to-one in 2007, then, the second one has been researched in grey demand under one-to-one in 2007, next, the dynamic cooperation mechanism in multi-stage trade under one-to-one has been studied in 2008, the dynamic cooperation mechanism under multi-to-one has been discussed in 2008. Based on above the researches, in this paper, The author is try to study the dynamic cooperation mechanism in supply chain for perishable agricultural products under one-to-multi. The grey matrix game model is established, the key factors to influence the cooperation between a farmer and many companies are analyzed. the dynamic cooperative incentive mechanism and measures to realize the united optimization of perishable products supply chain system are studies, and the dynamic cooperation mechanism to reach the performance of whole supply chain optimal is designed.

2 Assumptions and Notations

The market changes is assumed only two states: (i) the market price is higher than contract price; (ii) the market price is lower than contract price. Owing to the impact of natural disaster, market, national macro policies, and the future uncertain random fluctuation factors and so on, the market price is very strong uncertain: when it is higher than contract price or lower, the market price still changes in a interval. The number which only know its about scope and don't know it accurate value is a grey number^[14-18]. So, the market price of perishable agricultural products is assumed a interval grey number. The profits of "a farmer household & many companies" under contract price are still grey numbers. The game between a farmer household & many companies are grey game. The supply chain which is consist of a farmer and each company j is called channel j , In order to analyze the issue expediently, the following assumptions and notations are used throughout this paper: (1) In each supply chain channel j , the two players are rational economic men; (2) The market price is mutual knowledge; (3) There is the same the contract price in each supply chain channel j ; (4) There is the same penalty for company of an item in each supply chain channel j ; (5) There is the same bonus given to farmer of an item in channel j . In addition, the notations is shown in table.1:

Table 1. Variables notations

$P_{1m}(\otimes) \in [P_{1L}, P_{1H}]$ market price higher than contract price	Q_j the purchase quantity in channel j
$P_{2m}(\otimes) \in [P_{2L}, P_{2H}]$ market price lower than contract price	γ_j the bonus given to farmer of an item in channel j
C_f the cost of farmer	$\pi_{f1}(\otimes)$ the income of farmer under P in channel j
C_{cj} the cost of company j	$\pi_{c1}(\otimes)$ the income of company j under P
F_f the Penalty for farmer of an item	$\pi_{f2}(\otimes)$ the income of farmer under $P_{1m}(\otimes)$ in channel j
F_c the Penalty for company of an item	$\pi_{c2}(\otimes)$ the income of company under $P_{1m}(\otimes)$ in channel j
G_j the profit of company j	$\pi_{f3}(\otimes)$ the income of farmer under $P_{2m}(\otimes)$ in channel j
P the contract price	$\pi_{c3}(\otimes)$ the income of company under $P_{2m}(\otimes)$ in channel j
β_{ji} the competition influence coefficient between companies	θ the deterioration rate of perishable agricultural products

3 Grey Game Model and Analysis

3.1 Grey Game Model

The payment functions of company j and a farmer are their revenues .The income of farmer and company j under each situation:

(1) The revenues of farmer and company j under the contract price

$$\pi_{f1}(\otimes) \in [PQ_j - C_f - \theta PQ_j, PQ_j - C_f - \theta PQ_j + \gamma_j Q_j]$$

$$\pi_{c1}(\otimes) \in [G_j - (PQ_j + C_{cj} + \gamma_j Q_j), G_j - (PQ_j + C_{cj})]$$

(2) The revenues of farmer and company j under the market price which is higher than contract price

$$\pi_{f2}(\otimes) \in [P_{1L}Q_j - C_f - \theta P_{1L}Q_j - F_f Q_j,$$

$$P_{1H}Q_j - C_f - \theta P_{1H}Q_j - F_f Q_j]$$

$$\pi_{c2}(\otimes) \in [F_f Q_j, F_f Q_j]$$

(3) The revenues of farmer and company j under the market price which is lower than contract price

$$\begin{aligned} \pi_{f3}(\otimes) &\in [F_c(Q_j - \beta_{ji} Q_j), F_c(Q_j - \beta_{ji} Q_j)] \\ \pi_{c3}(\otimes) &\in [G_j - (P_{2H}(Q_j - \beta_{ji} Q_j) + C_{cj} + F_c(Q_i - \beta_{ji} Q_j)), \\ &G_j - (P_{2L}(Q_j - \beta_{ji} Q_j) + C_{cj} + F_c(Q_i - \beta_{ji} Q_j))] \end{aligned}$$

The payment matrix of company j and a farmer is a grey game matrix as figure 1.

		company [
farmer		performance	breach of faith
Performance	$\pi_{\square}(\otimes), \pi_{\square}(\otimes)$	$\pi_{\square}(\otimes), \pi_{\square}(\otimes)$	
Breachoffaith	$\pi_{\square}(\otimes), \pi_{\square}(\otimes)$	0, 0	

Fig. 1. The grey game matrix

The grey game matrix shows that when company j and farmer breach of faith at the same time, their profits are both zero. So, there are only three types:

- (i) When the market price is the same as the contract price , the corporation and the farmer cooperate. There revenues are $\pi_{f1}(\otimes), \pi_{c1}(\otimes)$ separately.
- (ii) When the market price is higher than contract price, the farmer breaches of faith, the farmer revenue is greater than the revenue when they cooperate, the company j revenue decrease. Their revenues are $\pi_{f2}(\otimes), \pi_{c2}(\otimes)$ separately, $\pi_{f2}(\otimes) > \pi_{f1}(\otimes), \pi_{c2}(\otimes) < \pi_{c1}(\otimes)$.
- (iii) When the market price is lower than contract price, the corporation breaches of faith, the company j revenue is greater than the revenue when they cooperate, the farmer revenue decrease. Their revenues are $\pi_{f3}(\otimes), \pi_{c3}(\otimes)$ separately, $\pi_{c3}(\otimes) > \pi_{c1}(\otimes), \pi_{f3}(\otimes) < \pi_{f1}(\otimes)$.

From above discussing , if the game between the company j and the farmer reach cooperation Nash equilibrium, the measures must be taken effectively to make their revenues are higher than these one under non-cooperation .That is

$$\pi_{f_1}(\otimes) \geq \pi_{f_2}(\otimes) \text{ and } \pi_{f_1}(\otimes) \geq \pi_{f_3}(\otimes),$$

$$\pi_{c_1}(\otimes) \geq \pi_{c_2}(\otimes) \text{ and } \pi_{c_1}(\otimes) \geq \pi_{c_3}(\otimes).$$

3.2 The Model Analysis

Theorem 1. when the market price is higher than the contract price, If,
 $F_f > (1 - \theta)(P_{1L} - P) - \gamma_j,$

$\gamma_j < (1 - \theta)(P_{1H} - P),$ then, $\pi_{f_2}(\otimes) > \pi_{f_1}(\otimes).$ The game result is that the farmer breaches of faith, and this is independent the competition between companies.

Proof. Let: $\pi_{f_2}(\otimes) \in [a_2, b_2] = [P_{1L}Q_j - C_f - \theta P_{1L}Q_j - F_fQ_j,$
 $P_{1H}Q_j - C_f - \theta P_{1H}Q_j - F_fQ_j]$
 $\pi_{f_1}(\otimes) \in [a_1, b_1] = [PQ_j - C_f - \theta PQ_j, PQ_j - C_f - \theta PQ_j + \gamma_jQ_j]$

So., we can get the Superiority Position Degree of the grey number $\pi_{f_2}(\otimes)$ relative

to the grey number $\pi_{f_1}(\otimes).$ $M_s = \frac{b_2 - b_1}{b_2 - a_2} = \frac{[(1 - \theta)(P_{1H} - P) - \gamma_j]Q_j}{(1 - \theta)(P_{1H} - P_{1L})Q_j},$

Since $(1 - \theta)(P_{1H} - P_{1L})Q_j > 0, \gamma_j < (1 - \theta)(P_{1H} - P), \Rightarrow M_s > 0,$ according to the grey system theory, the Inferior Position Degree of the grey number $\pi_{f_2}(\otimes)$ relative to the grey number $\pi_{f_1}(\otimes): M_l = 0, \Rightarrow M_s + M_l > 0.$

Moreover, we can have that $b_2 - b_1 = [(1 - \theta)(P_{1H} - P) - \gamma_j]Q_j > 0,$ It is easy to get that $b_2 > b_1.$ Similarly, we can get

that $b_1 - a_2 = Q_j\{F_f - [(1 - \theta)(P_{1L} - P) - \gamma_j]\} > 0, \Rightarrow b_1 > a_2,$ therefore

$b_1 > a_2.$ That is $b_2 > b_1 > a_2.$ Hence, we can conclude that

$\pi_{f_2}(\otimes) > \pi_{f_1}(\otimes)$ by the the grey system theory. The proof is therefore complete.

Theorem 2. When the market price is higher than the contract price, if
 $\gamma_j \geq (1 - \theta)(P_{1H} - P),$

$F_f < (1 - \theta)(P_{1H} - P),$ then, $\pi_{f_1}(\otimes) \geq \pi_{f_2}(\otimes).$ The game result is that a farmer cooperates with company $j.$ and this is independent the competition between companies.

Proof. The Superiority Position Degree of the grey number $\pi_{f_1}(\otimes)$ relative to the grey number $\pi_{f_2}(\otimes).$

$$M_s = \frac{b_1 - b_2}{b_1 - a_1} = \frac{[\gamma_j - (1 - \theta)(P_{1H} - P)]Q_j}{\gamma_j Q_j}, \text{ since, } \gamma_j \geq (1 - \theta)(P_{1H} - P),$$

$\Rightarrow M_s \geq 0$, according to the grey system theory, the Inferior Position Degree of the grey number $\pi_{f1}(\otimes)$ relative to the grey number $\pi_{f2}(\otimes)$: $M_I = 0$, $\Rightarrow M_s + M_I > 0$, Moreover, we can have that $b_1 - b_2 = [\gamma_j - (1 - \theta)(P_{1H} - P)]Q_j \geq 0$,

It is easy to get that $b_1 \geq b_2$, Similarly, $b_2 - a_1 = [(1 - \theta)(P_{1H} - P) - F_f]Q_j$, since, $F_f < (1 - \theta)(P_{1H} - P)$, $\Rightarrow b_2 \geq a_1$, therefore $b_1 \geq b_2 \geq a_1$, hence, $\Rightarrow \pi_{f1}(\otimes) > \pi_{f2}(\otimes)$, The proof is complete.

Theorem 3. When the market price is lower than the contract price, if $\frac{P}{1 - \beta_{ji}} - P_{2H} < F_c < \frac{P}{1 - \beta_{ji}} - P_{2L}$, then, $\pi_{c3}(\otimes) > \pi_{c1}(\otimes)$. The game result is that the companies breaches of faith.

Proof. Let,

$$\begin{aligned} \pi_{c3}(\otimes) \in [a_3, b_3] &= [G_j - (P_{2H}(Q_j - \beta_{ji}Q_j) + C_{cj} + F_c(Q_j - \beta_{ji}Q_j)), \\ &G_j - (P_{2L}(Q_j - \beta_{ji}Q_j) + C_{cj} + F_c(Q_j - \beta_{ji}Q_j))] \\ \pi_{c1}(\otimes) \in [a_4, b_4] &= [G_j - (PQ_j + C_{cj} + \gamma_j Q_j), G_j - (PQ_j + C_{cj})] \end{aligned}$$

so, we can get the superiority position degree of the grey number $\pi_{c3}(\otimes)$ relative to the grey number $\pi_{c1}(\otimes)$.

$$M_s = \frac{b_3 - b_4}{b_3 - a_3} = \frac{[P - (1 - \beta_{ji})P_{2L} - F_c(1 - \beta_{ji})]Q_j}{(1 - \beta_{ji})(P_{2H} - P_{2L})Q_j}, \text{ since,}$$

$$\frac{P}{1 - \beta_{ji}} - P_{2H} < F_c < \frac{P}{1 - \beta_{ji}} - P_{2L}, \text{ we can get that } M_s > 0, \text{ according to the}$$

grey system theory, the Inferior Position Degree of the grey number $\pi_{c3}(\otimes)$ relative to the grey number $\pi_{c1}(\otimes)$: $M_I = 0$, $\Rightarrow M_s + M_I > 0$.

Moreover, we can have that $b_3 - b_4 = [P - (1 - \beta_{ji})P_{2L} - F_c(1 - \beta_{ji})]Q_j > 0$, $\Rightarrow b_3 > b_4$.

then, $b_4 - a_3 = \{F_c(1 - \beta_{ji}) - [P - P_{2H}(1 - \beta_{ji})]\}Q_j > 0$, $\Rightarrow b_4 > a_3$, so, $\Rightarrow b_3 > b_4 > a_3$

Thus, according to the grey system theory, we can easy obtain, $\pi_{c3}(\otimes) > \pi_{c1}(\otimes)$, The proof is complete.

Theorem 4. If $\gamma_j \geq (1 - \theta)(P_{1H} - P)$,

$$F_f < (1 - \theta)(P_{1H} - P), \frac{P}{1 - \beta_{ji}} - P_{2L} \leq F_c \leq \frac{P}{1 - \beta_{ji}} - P_{2L} + \frac{\gamma_j}{1 - \beta_{ji}}$$

then, $\pi_{f1}(\otimes) \geq \pi_{f2}(\otimes)$, $\pi_{c1}(\otimes) \geq \pi_3(\otimes)$, The game result is that a farmer and the companies both cooperate.

Proof. Based on above the discussing in theorem 2, if $\gamma_j \geq (1 - \theta)(P_{1H} - P)$, and $F_f < (1 - \theta)(P_{1H} - P)$

Then $\pi_{f1}(\otimes) \geq \pi_{f2}(\otimes)$. The following proof is similar to theorem 3, since,

$$F_c \geq \frac{P}{1 - \beta_{ji}} - P_{2L}, \text{ so, we can easy get the superiority position degree of the grey}$$

number $\pi_{c1}(\otimes)$ relative to the grey number $\pi_{c3}(\otimes)$

$$M_S = \frac{b_4 - b_3}{b_4 - a_4} = \frac{\{F_c(1 - \beta_{ji}) - [P - (1 - \beta_{ji})P_{2L}]\}Q_j}{\gamma_j Q_j} \geq 0, \text{ according to the}$$

grey system theory, the Inferior Position Degree of the grey number $\pi_{c1}(\otimes)$ relative

to the grey number $\pi_{c3}(\otimes)$: $M_I = 0$, so, $M_S + M_I > 0$. Moreover, owing to

$$b_4 - b_3 = \{F_c(1 - \beta_{ji}) - [P - (1 - \beta_{ji})P_{2L}]\}Q_j \geq 0, \text{ so, } b_4 \geq b_3,$$

$$b_3 - a_4 = [P - (1 - \beta_{ji})P_{2L} + \gamma_j - F_c(1 - \beta_{ji})]Q_j \geq 0, \text{ so, } b_3 \geq a_4,$$

we can easy get $b_4 \geq b_3 \geq a_4$, based on the grey system theory, we can easy obtain

$$\pi_{c1}(\otimes) \geq \pi_{c3}(\otimes), \text{ The demonstration is finished.}$$

Theorem 5. When the market price is lower than the contract price, the competition between companies can reduce the default rates of companies, and the default rates of companies are decreased with the the competition influence coefficient between companies β_{ji} increasing.

Proof. When there is on competition between companies, if $F_c \geq P - P_{2L}$, com-

pany cooepetates, while $F_c \geq \frac{P}{1 - \beta_{ji}} - P_{2L}$, company cooepetates under competition.

Since, $0 < \beta_{ji} < 1$, so, $P < \frac{P}{1 - \beta_{ji}}$. Thus, we can obtain that the penalty for company of an item in competition is higher than the one in no-competition. the penalty for company of an item in competition is increased with β_{ji} increasing. So, the default rates of companies are decreased.

4 Numerical Example

The following example is illustrated to show the applicability and effectiveness of the above theorems .some vegetable process centers signs contract to a farmer , set $Q_j = 710$, $C_f = 200$, $C_{cj} = 100$, $P = 1.4$, $P_{1m}(\otimes) \in [3.2, 4.7]$, $P_{2m}(\otimes) \in [0.4, 0.7]$, $\theta = 0.02$, $F_f = 1.4$, $F_c = (1.4, 0.5, 0)$, $\gamma_j = (1.8, 1.0, 0)$, $\beta_{ji} = (0.8, 0.4, 0)$, $G_j = 5800$.

The various values of $\pi_{c1}(\otimes)$, $\pi_2(\otimes)$, $\pi_{c1}(\otimes)$, $\pi_{c3}(\otimes)$ which are calculated by using the software of MATLAB7.0 under different condition. and their variation trend are shown in figure 2, figure 3, figure 4.

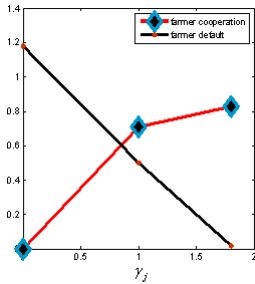


Fig. 2. The changing trend of a farmer cooperation/default of companies cooperation/default with γ_j changing

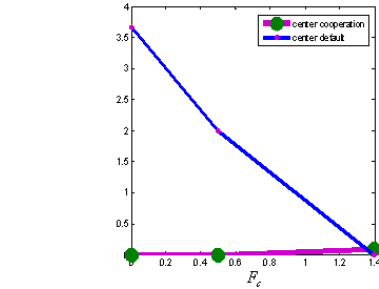


Fig. 3. The changing trend of with F_c changing

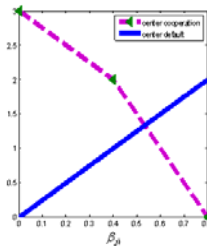


Fig. 4. The changing trend of companies cooperation/default with β_{ji} changing

From the figure 2, it is observed that when $\gamma_j < (1-\theta)(P_{1H} - P)$,and $F_f > (1-\theta)(P_{1L} - P) - \gamma_j$, then $\pi_{c2}(\otimes) > \pi_{c1}(\otimes)$, and the trend decreases with the increasing of γ_j . That is to say while the market price is higher than the contract price , the farmer income under market price is more than one under contract price, the farmer breaches of faith. As the bonus which vegetable process centers give to farmer increase, the farmer default rate reduces. From the figure 2 , it is also observed that when $\gamma_j \geq (1-\theta)(P_{1H} - P)$, and $F_f < (1-\theta)(P_{1H} - P)$, then $\pi_{c1}(\otimes) \geq \pi_{c2}(\otimes)$, and the trend increases with the increasing of γ_j .

Similarly ,from the figure 3, it is observed when $\frac{P}{1-\beta_{ji}} - P_{2H} < F_c < \frac{P}{1-\beta_{ji}} - P_{2L}$, then $\pi_{c3}(\otimes) > \pi_{c1}(\otimes)$. The companies breaches of faith and the trend reduces with the increasing of F_c . Moreover, its default rate reduces through increasing the punishment intensity. From the figure 3, it is also observed when $\frac{P}{1-\beta_{ji}} - P_{2L} \leq F_c \leq \frac{P}{1-\beta_{ji}} - P_{2L} + \frac{\gamma_j}{1-\beta_{ji}}$, then The companies cooperate and the trend of cooperation increases with the increasing of F_c .

From the figure 4, it is also easy observed that the competition between companies can reduce the default rates of companies, and the default rates of companies are decreased with the the competition influence coefficient between companies β_{ji} increasing.

5 Conclusion and Further Research

The above theorems imply that in the game Mode between a farmer and some companies, which compete each other. When the risk of breaching of faith is less and the mechanism of revenue sharing is not perfect, the senses of a farmer and some companies for performing contract are both weak. There are higher break contract rate ; If the responsibility of bearing risk reinforces, the effective revenue sharing is set up between a farmer and some companies, and the effective dynamic cooperation mechanism is established to make the incomes of a farmer and some companies performing contract must not be less than that of breaking contract and the a farmer and some companies are effective supervised . Those are the optimization equilibrium strategy and measure to maximize both sides profit in two- stage perishable agricultural products supply chain system. The results also show that the competition between companies can reduce the default rates of companies, and the default rates of companies are decreased with the the competition influence coefficient between companies increasing. It is independent the competition of companies that the farmer cooperates or breaches of faith.

This paper applies the grey system theory to supply chain management and researches the dynamic cooperation incentive mechanism for perishable agricultural

products supply chain system. It is a new try for supply chain management to use grey system theory either in theory or in practice. Our next study is to research grey game in agricultural perishable products supply chain under multi-to-multi.

References

1. Chen, C.T., Lin, C.T., Huang, S.F.: A Fuzzy Approach for Supplier Evaluation and Selection in Supply Chain Management. *International Journal of Production Economics* 2, 289–301 (2006)
2. Guo, M., Wang, H.W.: The Coordinative and Incentive Mechanism in Cooperative Supply Chain. *Systems Engineering* 4, 49–53 (2002)
3. Guo, M., Wang, H.W.: Qu T. Incentive Mechanism Analysis in 2 - Stage Stochastic Supply Chain. *Computer Integrated Manufacturing Systems* 12, 965–969 (2002)
4. Xie, S.Y.: Evolutionary Game Theory Under Bounded Rationality. *Journal of Shanghai University of Finance and Economics* 3, 3–9 (2001)
5. Nagarajan, M., Sosis, G.: Game-theoretic Analysis of Cooperation among Supply Chain Agents. *European Journal of Operational Research* 5, 1–27 (2006)
6. Simon, R.S.: The Structure of Non-zero-sum Stochastic Games. *Advances in Applied Mathematics* 38, 1–26 (2007)
7. Fu, H.: Pure Strategy Equilibrium in Games with Private and Public Information. *Journal of Mathematical Economics* 43, 523–531 (2007)
8. Smirnov, A.V., Sheremetov, L.B.: Soft-computing Technologies for Configuration of Cooperative Supply Chain. *Applied Soft Computing* 4, 87–107 (2004)
9. Bylka, S.: Competitive and Cooperative Policies for the Vendor–buyer System. *International Journal of Production Economics* 81, 533–544 (2003)
10. Nie, P.Y., La, M.Y., Zhu, S.J.: Dynamic Feedback Stackelberg Games with Non-unique Solutions. *Nonlinear Analysis: Theory, Methods & Application* 69, 1904–1913 (2008)
11. Jia, O.J.: Research on Cooperation Mechanism of 'Company & Peasant Household' Mode Based on Analysis of Feed-back Dynamic Complexity. Doctoral Dissertation Nanchang University 5, 50–60 (2006)
12. You, C.M.: The Contract Design and Price Mechanism of 'Company & Peasant Household'. *Economic Problems* 2, 57–58 (2004)
13. Sun, Y.W., Liu, Z.: Economic Analysis on Operation Difficulties of "Corporation & Farmer" Organization. *The Theory and Practice of Finance and Economics* 25, 113–118 (2004)
14. Liu, S.F., Dang, Y.G., Fang, Z.G.: *The Grey System Theory and Application*. Science Press, Beijing (2004)
15. Fang, Z.G., Liu, S.F.: Grey Matrix Game Model Based on Pure Strategy. *Journal of Nanjing University of Aeronautics & Astronautics* 35, 441–445 (2003)
16. Mi, C.M., Fang, Z.G.: Study on Strategy Dominance and Pure Strategies Solution of Grey Matrix Game Based on Interval Grey Number Not to Be Determined Directly. *Chinese Journal of Management Science* 13, 81–85 (2005)
17. Fang, Z.G., Liu, S.F., Ruan, A.Q.: Pure Strategy Solution and Venture Problem of Grey Matrix Game Based on Undeterminable Directly Interval Grey Number. *Journal of Jilin University* 36, 137–142 (2006)
18. Wang, L.J., Wang, H.W., Sun, X.C.: Research Cooperate-game for Agricultural Perishable Products by Using Grey Theory. *Journal Huazhong University of Science and Technology* 8, 30–33 (2008)

Primary Research on Urban Mass Panic Based on Computational Methods for Experiments

Xi Chen, Qi Fei, and Wei Li*

Institute of Systems Engineering, Huazhong University of Science and Technology,
Key Laboratory of Ministry of Education for Image Processing and Intelligent Control
430074 Wuhan, China
shirlyli_hust@126.com

Abstract. Base on the theory of computational methods of experiments and agent modeling technology, combining with sociology and psychology, the paper presents method of computational experiments for research of Urban Mass Panic (UMP). The method proposed a panic-information transmitting artificial system for computational experiments with agent modeling technology. The system planed two kinds of agents: social individual agents and panic alarming agents. Social individual agents are designed to simulate persons living common human society. These agents which contain the ability of panic information recognition, reflection and making different decision are simulated with some nonlinear methods, such as fuzzy neural network. Panic alarming agents are designed to simulate circumstance which engenders panic information or panic phenomenon. These agents collect information about panic and release information to social individual agents by different transmitting methods we can image. The paper also discusses how to carry through measurement to check the level of UMP by the information entropy theory dynamically.

Keywords: Computational Methods for Experiments (CME) , Urban Mass Panic (UMP).

1 Introduction

The Urban Mass Panic (UMP) is always defined as people's collective mental reactions and behavioral reactions which were noncooperation and inconsequence, because of people's ignorance of these events they faced [1].The water crisis, in Jilin City starting November 22,2005, is a classical sample for UMP: A massive explosion took place in a petrochemical plant in neighboring the city on November 13. With a government announcement of a four-day stoppage of water supply after 9 days, many rumors about earthquake, terrorism attack and water pollution spread quickly in the city. These panic phenomena, such as buying water and escaping from this city, emerged gradually. Therefore, it is very import to research the characters and mechanisms of UMP caused by emergencies because UMP is a kind of objective existence. The traditional research

* Corresponding author.

of UMP always adopted methods which gather and analysis the information of reactions of human being after or when the emergency took place (such as social alarming system[2], a web mining based measurement and monitoring model of UMP [3]). Those methods can only be used to analyze the situation which has lead to serious panic result and cannot be used to present developing and spreading rules of UMP clearly from the point of view of these essence characters and internal mechanisms.

Scholars, in the field of systems engineering, have presented the Computational Methods for Experiments (CME) or Computational Experiments (CE) [4,5] for re-researching complex social systems. Those scholars wanted to become the CME to serve as a laboratory replacing with these actual nature systems and do different experiments related to system’s actions or decision analysis. It provides a new thought to study the UMP phenomena.

A computational method for experiments is given in this paper. An agent-base panic-information transmitting artificial system, which study these actions of panic information recognition, reflection and making different decision, has been designed. At last, the way to measure the level of UMP dynamically by the theory of information entropy is also discussed.

The rest of the paper is organized as follows. Section 2 discusses related works in field of UMP and CME. Section 3 presents the design of agent-base panic-information transmitting artificial system and the measurement for the level of UMP by the theory of information entropy. Section 4 concludes.

2 Related Works

2.1 Urban Mass Panic’s Origin

There are many reasons which lead to UMP, but tale or rumors’ spreading may be the most significant reason. Many scholars in the field of sociology and psychology have got many achievements from studying rumor and its spreading. The latest definition of the rumors is that statements or annotations about those issues concerned with public, which are transmitted by means of public or non-public methods and unproved or not based on definite knowledge[6]. There were many statements about the generating mode for rumor and these most popular ones are given behind. Allport defined the mode as[7]:

$$R = i \times a \tag{1}$$

- R:rumor
- i: importance. The importance of the issue involved;
- a :ambiguity. The ambiguity of the issue involved.
- Cross add a parameter- c to the mode ^[8]:

$$R = i \times a \times c \tag{2}$$

- c: the critical ability of audience.

Cao and Huang proposed that the circumstance in which rumors were transmitted is also very important[6]. They deemed that any rumor’s transmission cannot success

without these three behind factors: transmitter, receiver and the circumstance of rumor's transmission.

Recently, many scholars have focused on the UMP issue- public psychology problems and inadaptable actions which were largely due to the rumor's transmission in public emergencies. Liu has studied the process people how to cognize and react the emergency originally under the restriction of outside circumstance and individual factors[9]. Wang has studied two methodologies of psychology to social alarming system: The relation between attitudes and reactions, sampling investigation issues[4]. As earthquake took place frequently and always had great influence, the earthquake rumor has been studied more deeply. Gu has studied the cause of the earthquake rumor [10]. He presented that it is the most essential cause that the earthquake forecast hadn't been hold by human being, it is the recognition source that the lack of earthquake knowledge and it is the psychology factor that the horror of earthquake disaster. Hong has studied the issue why earthquake rumor transmitted more and more quickly and become more and more strange[11]. These research works listed above almost studied rumors by mathematical induction methods, such as statistical analysis, from the point of view of sociology and psychology and based on those phenomena or data after or when the emergency took place. These research works listed above hadn't formed theories and methods of experiments, which adopt these latest information technologies, such as computer simulation or data fusion, to research UMP's phenomenon and mechanisms. Li and Chen had tried to analyze UMP's characters and help to seize panic's spreading[3]. They wanted to implement a kind of online analysis method which adopt OLAP(On-Line Analytical Processing) to do web-based investigation including discovery, measurement and monitoring. However these available research works about UMP have some shortages, especially in the field of the internal mechanisms of UMP caused by rumors spreading and transmitting rapidly in different information transmitting networks. On the whole, if we take a careful consideration, it is not difficult to draw the conclusion that researchers should search more effective research techniques to analyze the internal mechanisms and to measure complex horror information generated from people society.

2.2 Computational Methods for Experiments

At present, computer simulation has entered a new era when parallel computing, distributed computing and grid computing, and other large-scale simulation methods and structures become common. Simulation has developed quickly, from simulating the natural process based on the differential equation to simulating the artificial or combination process of social system and human behaviors. Because of inherent subjectivity of the complex system, such as human or society, we can further regard the "simulation" results as a realistic alternative version, or a possible reality, and only regard the actual system as a possible reality, just like a kind of simulation result [4]. This thought is the thinking foundation which changes from the computation simulation to the computational experiments.

In the computational experiments, the traditional computation simulation has become a "test" process in "computation lab ", and become an approach to "grow or train" all kinds of complex systems. At the same time, the actual system only becomes a result in the "computational experiments". As a result, computational experiments are

different from the common computation simulation, because the common computation is compliance with the belief which the actual system is the only real existence and it is the only aim to simulate the actual system realistically and it regards the actual system as the only reference standard to test the success of the simulation and to seek the "truth". In the computational experiments, the computation simulation is also a kind of "reality" and may be a substitute form for the actual system or another possible way to achieve [4]. According to this understanding, we can carry out "computational experiments" by computation simulation naturally.

Many scholars had done some in-depth studies and applications, since the methodology of the computational experiments was presented. Wang studied the framework, main approaches and problems of the computational experiments[5]. Wang also analyzed its status in the field of complex social-economic system. Miao and TANG researched and implemented the artificial transport system based on the computational experiments[12]. Cai and Zhao studied combat system with the computational experiments[13]. From these researches and applications, the computational experiments is beneficial to change the situation in which the study of complex systems rely on the induction and the deduction too much simply and to provide a kind of method of induction- deduction based on computer simulation experiments with the adoption of integrated systems science, management science and information science. The new method will provide advantages to in-depth research of complex system, such as human society.

On the whole, in the field of public safety and emergency management, the existing researches are short of advanced theories and approaches of experiments to study and analyze the internal mechanisms of UMP in human society. At the same time, in the field of complex systems, the method of computational experiments is rising day by day. At present, the research works which combine with the tow fields and promote each other haven't been started, that is what this paper wants to discuss.

3 Research Thought of Computational Methods for Experiments for UMP

The approach of computational methods for experiments for UMP involves three parts: Firstly, the approach needs to study the model and specification of UMP problems using knowledge of sociology, psychology and journalism. Secondly, it needs to design an agent-base panic-information transmitting artificial system. At last, it needs to design a measurement model-UMP measurement model to measure the panic level. Figure 1 shows the framework of computational methods for experiments for UMP. The agent-base panic-information transmitting artificial system, which contains the first and second steps, is responsible for receiving information from external environment (such as emergency information, networks environmental information), and simulates the process of individual panic information's transmission by instantiating, starting and stopping agent-related entities, according to the external environmental information. The artificial system also is responsible for collecting information of interaction between different agents and generating experimental data. UMP measurement model is responsible for analyzing information from agent-base panic-information transmitting artificial system and measure or evaluate degree of public panic dynamically and continuously.

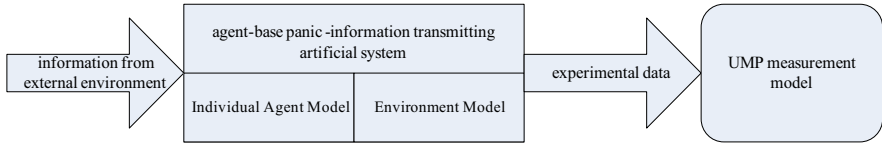


Fig. 1. The actual framework of computational methods for experiments for UMP

3.1 Agent-Base Panic-Information Transmitting Artificial System

Agents are the autonomous programs situated within an environment (either a host or a network), which sense the environment and acts upon it to achieve their goals[14,15].An agent uses the internal world state information and inference engine to compute the actions to be performed on the environment either by sensing environment or upon reception of messages from other agents/users[16]. Multi-agent systems contain a group of agents which have certain resource and capability, relatively independent and interactive cooperation. As the multi-agent systems provide a multi-level bottom-up approach which research complex systems, it can offer fundament effectively for studying complex systems -such as human society. As a result, we can establish the agent-base panic-information transmitting artificial system for researching panic mechanism. This artificial system contains individual agent model and environment model. The architecture of the artificial system is shown in Figure 2 as below:

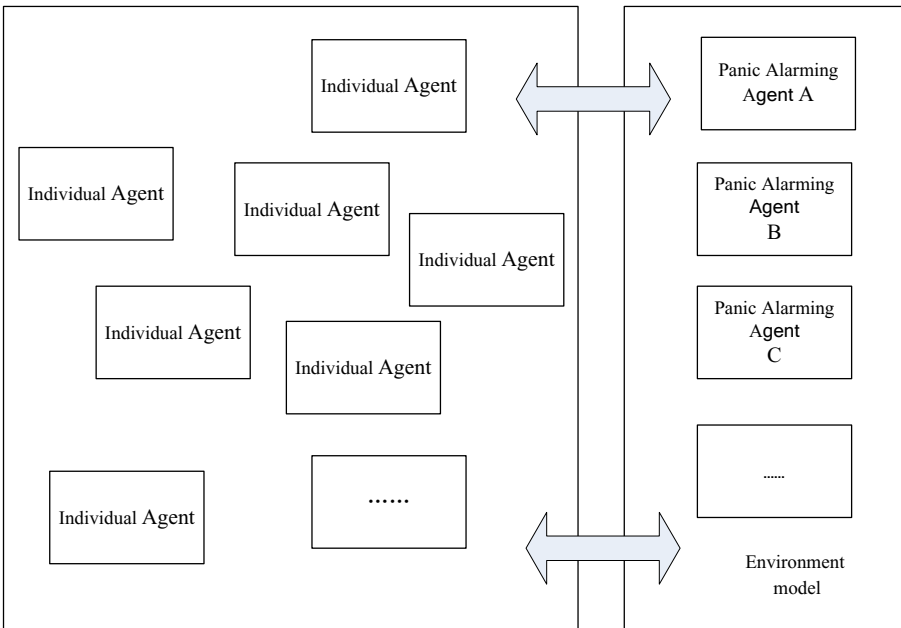


Fig. 2. Architecture of agent-base panic-information transmitting artificial system

The environment model is implemented by a number of panic alarming agents. The panic alarming agent is, actually, abstract models for external crisis alarming system and it can present some units which have the capacity to control or produce the information's broadcast, such as governments, media units. The following contents take earthquake alarming information as example to describe the process of panic alarming agent how to react the warning information which maybe produce rumor. Receiving the earthquake warning or alarming information, these panic alarming agents' judge by the rules in the rule bank, which can be maintained dynamically, and draw a conclusion that whether and how to convey this warning message to all social agent entities. Warning or alarming information contains the warning information's classification, occurrence time, place, agent receiving information and warning intensity and so on.

At present, for the individual in society, the UMP caused by public emergencies involves people's physical and psychological issues, and the public panic information involves features of transmitting networks. Therefore, it is necessary to research public panic information's classification, society factors and individual factors from the point of view of panic information's interactions. At the same time, it also need to study networks where panic information transmits and the networks' characters based on these theories of information transmission and complex network. In addition, the process from cognize behaviors to decision-making also should be reflected in agent's model.

In case of great many inaccuracy, imperfection and uncertainty of panic information received from other social individual agents, the information often cannot be describe as precise "true" or "false". As a result, it is impossible to use linear methods which need precise data when we need to simulate how to a large number of external environment panic information which may be conflict with each other can be accepted by social individual and how to produce the relevant response. As the fuzzy theory being beneficial to describing the inaccuracy, imperfection and uncertainty, it is possible, in the framework of fuzzy theory, to fuse panic information which come from different source and information transmitting networks, and then to simulate people react to panic information by reasoning and judging methods, at last to simulate the spread of panic information.

3.2 UMP Measurement Approach Based on Information Entropy Theory

It would be unlikely to be able to adopt one certain common analysis evaluation method to measure public panic information related to many different factors. At present, conventional analysis evaluation methods involves statistical analysis, AHP (Analytic Hierarchy Process), data mining and so on. Based on perfected theory and mathematical skill, Most of the statistical analysis techniques' forecast accuracy is satisfying, but it is high demand for users. Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. It is the process of finding correlations or patterns among dozens of fields in large relational databases. But, data mining need a large amount of historical data-it is sometimes so difficult. AHP is a structured technique for helping people deal with complex decisions. AHP is clear, easy to user and also needn't a larger amount of data. However, all

evaluation must be based on the same weight system. It is difficult to confirm a standard weight system in actual situation which contains various factors (time, place, occupation, etc.). Therefore, we present a kind of dynamic measurement approach for measuring public panic information. Based on the theory of information entropy, the approach integrates statistical analysis method and clustering technique in data mining.

Concrete method can be summarized as follows: 1) Design rules (can be changed dynamically) of data mining, according to data related to each other (social environmental and individual factors, such as age, occupation, sex or spreading networks, etc.). 2) Carry out data mining for huge data getting from computational experiments with clustering technology. Then do some data collation, explore data relationship and make a reasonable evaluation to these different types of earthquake panic information which produce from data mining. At last, help researchers or officers to find different characteristics groups in earthquake panic information. 3) Combine with information entropy theory, according to different characteristics groups in earthquake panic information as well as definition and classification of the earthquake panic information, calculated different information entropy of panic related to certain characteristics or certain levels (such as general information entropy of earthquake panic, dangerous level information entropy of earthquake panic, female information entropy of earthquake panic in certain region etc.). As result, we can measure total quantity of earthquake panic information caused by relative characteristics' information sources. In addition, we can define total quantity of interactive information getting from computational experiments as total information entropy of panic, and define the ratio of different characteristics' earthquake panic information and total information entropy of panic as the ratio of different characteristics' earthquake panic information. These ratios can provide a wide range of research data and samples.

4 Conclusion

UMP not only is an old problem accompanied with human society existing, but also is a new thing keeping on evolving and developing. With the development of human society and science, UMP's research becomes further and improving. This paper presents a new way to study UMP, which combine with sociology, psychology and computational methods of experiments in the field of computer simulation. Of course, this new approach just a research prototype. It is a long way to go on the way. We will do further research in the field in the future.

Acknowledgments. The authors are grateful to all participators for their valuable comments and suggestions that have lead to the improvements of this paper. The first author is also grateful to his superior for supporting this research. Great thanks for the support from National Natural Science Foundation of China (NSFC, Grant 70671045).

References

1. David, G.M.: *Social Psychology*. McGraw-Hill, New York (1993)
2. Wang, E., Zhang, B.B.: The Role of Psychology to Social Alarming System. *Advances in Psychological Science* 11, 363–367 (2003)

3. Li, M.L., Chen, A.: A Web Mining Based Measurement and Monitoring Model of Urban Mass Panic in Emergency Management. In: Fifth International Conference on Fuzzy Systems and Knowledge Discovery, pp. 366–370. IEEE Press, New York (2008)
4. Wang, F.Y.: Computational Experiments for Behavior Analysis and Decision Evaluation of Complex Systems. *Journal of System Simulation* 16, 893–897 (2004)
5. Wang, F.Y.: Artificial Societies, Computational Experiments, and Parallel Systems: A Discussion on Computational Theory of Complex Social-Economic Systems. *Complex Systems and Complexity Science* 1, 25–35 (2004)
6. Cao, N.P., Huang, X.: Research on the Phenomenon of Rumor in Network Transmit. *Information Studies. Theory & Application* 27, 586–589 (2004)
7. Allport, G.W., Postman, L.: *The Psychology of Rumor*. Henry Holt, New York (1947)
8. Lowery, S.A., DeFleur, M.L.: *Milestones in Mass Communication Research*. Allyn & Bacon, Needham Heights (1995)
9. Liu, Y., Chen, A.: Research on the Psychology Reaction Mechanism in Emergency Management. *Modern Business Trade Industry* 20, 59–60 (2008)
10. Gu, F.Q.: The Root of Earthquake Rumors and Some Principles of Countermeasure. *Journal of Catastrophology* 3, 33–37 (1988)
11. Hong, S.Z.: Some Issues on Stopping Earthquake Rumors. *SiChuan Earthquake* 1, 31–37 (1983)
12. Miao, Q.H., Tang, S., Wang, F.Y.: Design of Artificial Transportation System Based on JXTA. *Journal of Transportation Systems Engineering and Information Technology* 6, 83–90 (2006)
13. Cai, Y.F., Zeng, X.Z.: Research Method of Complex Combat Systems Based on Computational Experiments. *Journal of System Simulation* 17, 2239–2243 (2005)
14. Franklin, S., Graser, A.: Is it an Agent or Just a Program. In: *Proceedings of the International Workshop on Agent Theories*, <http://citeseer.nj.nec.com/32780.html>
15. Bradshaw, J.: *Software Agents*. AAAI Press, California (2000)
16. Magedanz, T., Rothermel, K., Krause, S.: Intelligent Agents: an Emerging Technology for Next Generation Telecommunications. In: *Fifteenth Annual Joint Conference of the IEEE Computer Societies*, pp. 464–472. IEEE Press, New York (1996)

Virtual Reality Based Nuclear Steam Generator Ageing and Life Management Systems

Yajin Liu¹, Jiang Guo¹, Peng Liu², Lin Zhou², and Jin Jiang¹

¹ School of power & mechanical engineering, Wuhan University,
Wuhan 430072, China

² China Guangdong Nuclear Power Group, Shenzhen 510008, China

Abstract. Steam Generator (SG) is the barrier between primary loop and second loop in a nuclear power plant. As an important device, once SG fails to work due to the aging problem, the safety, economic and reliability of the nuclear power plant will be affected seriously. An important issue in the ageing and life management of SG is the understanding of its structure, operation mechanism and its ageing mechanism. Virtual reality is an advanced human-computer interface that simulates a realistic environment. Virtual environment technologies have been shown to provide invaluable training in the performance of complex procedural tasks. Of paramount importance to the success of any VR is realism; entities must move and behave believably and approximate to the complexity of the real world. A PC (Personal Computer)-based visualization system of SG is developed. The physical structures and the internal interactions of SG in various conditions, even those are invisible in the real world, can be visually demonstrated in detail by this system. This development system can provide vital indications of the equipments' ageing condition, integral evaluation of SG tube and also be useful for the staff training in the Nuclear Power Plants (NPPs).□

Keywords: Virtual Reality (VR), Steam Generator (SG), Ageing and life management, Nuclear power plants (NPPs), Training.

1 Introduction

VR systems use software and hardware to create and manage a virtual, interactive 3D environment that includes visual and sometimes audio and tactile elements. They generally include various types of display, sensor, and user-tracking and -navigation technologies. The systems can either simulate a real environment, such as a building, or create an imaginary one. With this system, participants can cruise around in the virtual world, behave believably and approximate to the complexity of the real world. They can see the object from different angles and also reach into it, then grab and reshape it. Neither symbol on screen for manipulation nor commands is required to run the program in the computer. As the VR technology has the advantages shown above, it offers accessible and cost-effective means for training personnel who currently made little or no experiential preparation for their assigned task.

In recent years, virtual reality has benefited tremendously from a variety of fields. Its implementation involves computer graphics advances, ranging from computing platforms, display technology, and software techniques to the understanding of human perception. The results of this progress are evident in the increasing number of VR-based training systems, immersive visualization systems, and visually realistic video games.

The nuclear power plant (NPP) is a very complicated structural system that is of strict requirements for the nuclear safety. And the nuclear safety is a matter of primary importance for a NPP. As a barrier between primary loop and second loop in a NPP, once SG fails to work, the safety, economic and reliability of the NPP will be affected seriously. Moreover, the public health and safety will be damaged, which is more serious. The ageing of SG has been recognized for many decades as an important cause of the premature failure of a wide variety of NPPs in worldwide scale. Consequently it is a topic that has received considerable attention. Much effort has been expended in attempting to understand the processes of ageing and degradation, what types of test should be undertaken to ascertain the condition of SG, and how short-term accelerated testing might be carried out in order to forecast the long-term life of SG.

At the end of 2008, China has 11 reactors in commercial operation. Table 1 is the information of Chinese mainland NPP steam generator. By the year 2014, the two main NPPs, QNPC and DayaBay will have been operating for twenty years. There is, nonetheless, a general desire to continue operating them for as long as possible. As the population of commercial nuclear power plants has matured, the need to manage the SG ageing problems and extend the SG life has increased.

Table 1. Information of Chinese mainland NPP steam generator

NPP	Reactor type	SG type	SG number	Commercial operation
QNPC	PWR	300MWe	1×2	1994.4
NPQJVC	PWR	60F	2×2	2002.4
				2004.5
TQNPC	CANDU	Candu-6	2×4	2002.12
				2003.7
DayaBay	PWR	55-19B	2×3	1994.2
				1994.5
Lin Ao	PWR	55-19B	2×3	2002.5
				2003.1
TianWan	WWWER-1000	PGV-1000	2×4	2007.5
				2007.10

Within the past 5 years SG Ageing and Life Management (ALM) have become important facets of nuclear power plant operations in China. Related researches are as follows:

- **QNPC**
2003, AMP & SGAMDB development
2007, SG lifetime evaluation
- **TQNPC**
2005, ageing mechanism analysis, AMP & SGAMDB development
- **DayaBay**
System and equipment screening, SGAMDB development
- **TianWan**
SG Ageing and life management

2 ALM-VR System

The technique of SG ALM has been fruitfully developed since 80s of the last century. Results are obtained from the research. Some of the results have been applied to the ageing evaluation of the SG. However, quite large numbers of them are not as expected. Most of the ageing location and life evaluation have to be made by human experts. An ALM system for the operation of SG is proposed in this paper. The architecture of ALM system is shown in Fig. 1.

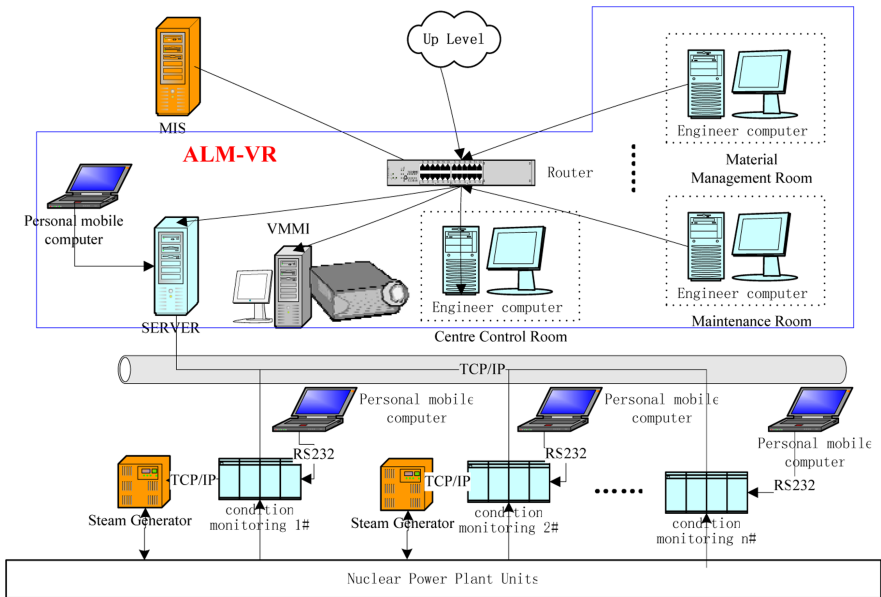


Fig. 1. ALM system architecture

It is primarily important to collect information for ageing management. Three approaches can be used for the information collection about the condition of the equipment: on-line continuous condition monitoring; on-line or off-line periodic test; and human inspection. Terabytes of historical data are stored in the Server. It consists a great

deal of experience obtained from other plants, such as the ageing problems observed; the location and type of ageing that has occurred; the date that ageing problem was first detected and the progression rate of that problem; how sleeving, peening, temperature reduction, or heat treatment may assist in long term SG operation and so on.

The ALM system will be externally integrated with the VMMI and the MIS to perform the simulation of the mechanism of the human society and to achieve higher performance and reliability.

Details of the system ALM-VR are described below.

2.1 Structure of the ALM-VR System

ALM-VR is realized as a part of the ALM system, which is mainly designed to offer the ageing condition indication, the interactive evaluation, the maintenance guidance and the staff training. It is based on low-cost PCs. The advantages of the developed system are as reducing the cost of the VR software itself and preserving realism and real-time performance. The system structure of ALM-VR is shown in Fig.2.

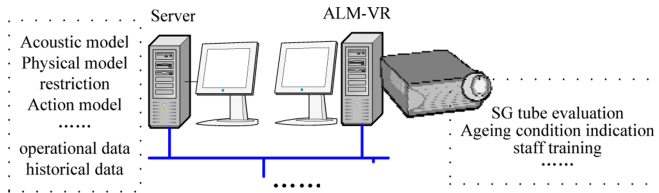


Fig. 2. ALM-VR hardware structure

The Server is to store information of the whole system such as the historical data of SG, the operational data of SG, the fault/detect information from other plants etc. Additional information such as physical model, action model, acoustic model etc. used for visible normal and fault feature simulation are also stored in the server.

ALM-VR is an advanced human-computer interface. It provides vital indications of the condition for the equipment, warnings of latent ageing problems, feedback experience, and online guidance. As a result, it is also a useful tool for the training of the maintenance engineers.

Human experts can log in the ALM-VR Website, where they can acquire the information of the equipment easily. Also they can exchange the experiences of the maintenance with the system. Upon the occurrence of the degradation/fault/detect of equipment, human experts can on-line participate in diagnosis and make the final judgments. The experts even need not to attend the spot and can solve problem remotely.

The operators can also acquire the equipment condition through the internet instead of on duty.

2.2 Software Tools

While a comprehensive coverage of currently existed VR development software on market or available for public use cannot be performed here, a brief survey of popular

VR toolkits or software function libraries for VR application development is presented here to situate ALM-VR within the context of existing VR software. It integrates different software and toolkits to realize a virtual environment. It involves Actify SpinFire, SolidWorks and so on.

Actify SpinFire™ Professional enables manufacturing organizations and their supply chains to easily access, interact with, and communicate part data plus related files and documentation. All major 3D and 2D CAD data formats allowed. SpinFire Professional streamlines the communication of CAD files and related data by enabling users to save and share this design data as a compact .3D file. It cuts costs and boosts productivity across the manufacturing enterprise and supply chain by enabling faster, more efficient communication of 3D and 2D design data. Use it to produce quick and accurate quotes, inspect the quality of 3D designs while on the shop floor to limit excess scrap, and much more.

However, appropriate modeling of the equipment is remained a problem to be solved. Although Actify SpinFire is very apt to realize various kinds of the model varying in color, illumination, lamination, mutual operation and cartoon, it only offers basic geometry modeling function of element, and makes complicated modeling relative difficulty of model.

SolidWorks on the other hand provides an effective way to solve this problem. It is the world™s best-selling professional 3D modeling, animation and rendering software for creating visual effects, character animation and next generation game development. SolidWorks delivers a fully collaborative 3D environment and new high-speed interactive rendering. Its completely customizable and extensible architecture allows for absolute artistic freedom.

3 Implementation of the ALM-VR

ALM-VR is an advanced graphic user interfaces (GUIs). It consists of following four levels, which provide information of SG to the operators in NPPs: (1) 3D model structure of SG; (2) vital 3D indications of the SG's condition; (3) integral evaluation of SG tube and (4) staff training. Details of the four levels are given below.

3.1 3D Model Structure of SG

SG contains many complexity components. These components work together to transfer heat from the primary system to the secondary system. To understand components of the SG and its operation is not an easy task for many engineers and managers, especially their mechanical interactions. In conventional methods, the main structure information about the SG is displayed with 2D graphics. It is difficult for operators to effectively master the structure information. For this reason the ALM-AR changes such a traditional representation. It utilizes three-dimensional model technique to develop the visualization system. In this system, the user can freely control the viewing angle of the object as if they move the object with the feeling of total immersion in the environment. The user can understand the basic information of SG and disassemble it. The visualization of upper internals is presented in Fig.3.

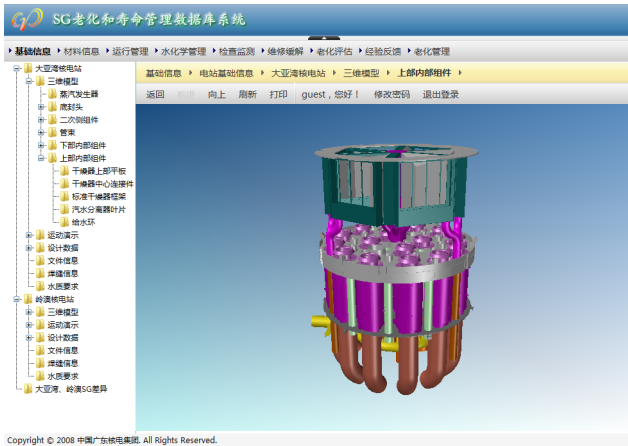


Fig. 3. The visualization of upper internals

3.2 Vital 3D Indications of the SG’s Condition

SG contains many different subsystems. Each of the components has to be properly specified, tracked, and inspected. This needs to create literally millions of documents. In conventional methods, the main information about SG is displayed with forms of figures and numbers. It is sometimes difficult for operators to effectively handle such information especially in the urgent condition. For this reason ALM-VR changes such a traditional display method. It utilizes three-dimensional model technique to develop the virtual environment system in which the operators can understand the basic information of the equipment and find out the relationship of the information with corresponding equipment. Fig.4. is the three-dimensional indication of SG.



Fig. 4. 3D indication of SG

3.3 Integral Evaluation of SG Tube

Once SG fails to work, the safety, economic and reliability of the NPP will be affected seriously. Moreover, the public health and safety will be damaged, which is more serious. As the central component that transfers heat from primary side coolant to secondary side water, the SG tube which is only about 1mm thick is a weak and pivotal element of SG. And its condition is a matter of primary importance for a steam generator. The historical overview of all tube repairs for a SG (Fig.5) provided by the ALM-VR gives the benefits of extending equipment lifetimes, reducing risk of failures and avoiding significant safety incident.

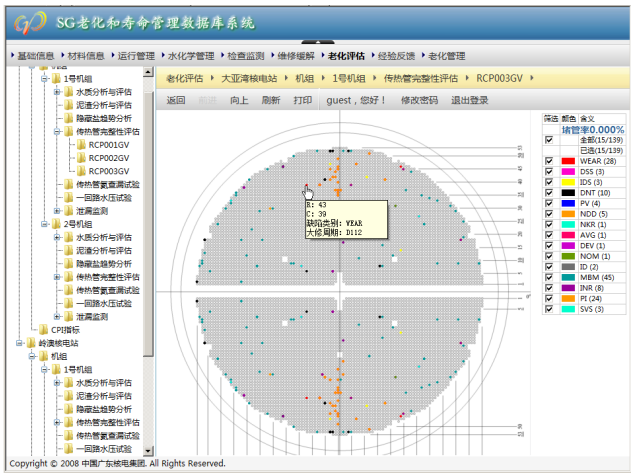


Fig. 5. The historical overview of all tube repairs for a SG



Fig. 6. The historical overview of a particular tube repairs

This window displays all the defect tubes of this SG grouped by defect style. The simple defect information can be got when the mouse approach a particular tube. By clicking it, the tube scene is loaded in the main window and its related information is also displayed (Fig.6). By this way, users can easily get the defect location and other detail information of the tube.

3.4 Staff Training

Staff-training programs for NPPs must take into account high safety requirements. These programs must efficiently train workers to manage both normal and abnormal operational conditions and to respond correctly to abnormal conditions through established procedures.

In the past, staff training employed physical copies of NPP's subsystems (for example, the reactor core and the instrumentation and control systems), with the same structure as the real ones. Because these copies were large and expensive, computer-based NPP simulation has become an alternative. ALM-VR is an advanced computer-assisted training system using VR technology. Compared with the traditional training approaches, the developed system offers accessible and cost-effective means for training personnel who currently made little or no experiential preparation for their assigned task and also allows the trainees to properly operate new equipment before its actual installation. Besides demonstrating the installation, operation and the maintenance procedures, this system can also simulate incidents or accidents. In addition to these advantages, ALM-VR provides all of the electronic documents of SG, such as the 3D model structure of SG (Fig.3), the welding information, the water quality requests, design documents etc. Fig.7 is the welding information of SG, including the detail information of the welding and the 2D graphic.

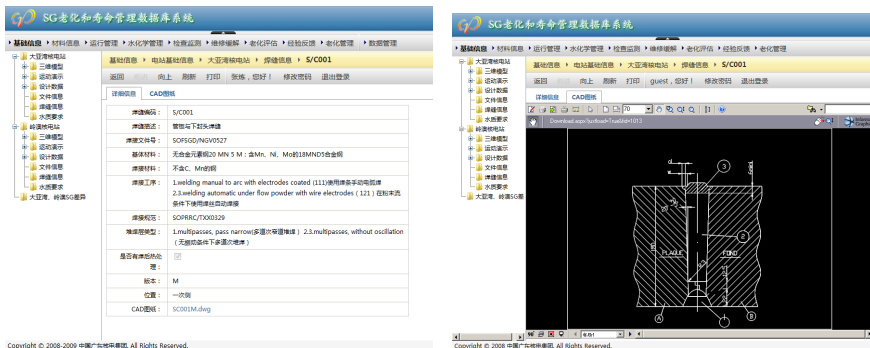


Fig. 7. The welding information of SG

The important perceptual cues and multi-modal feedback (e.g., visual, auditory and haptic) provides to the trainees the capability to effectively transfer from the virtual training to the real-world operation skills. More importantly, as the system provides higher freedom to the operators and the results of improper operation can be simulated, the additional injure of human and equipment can be avoided.

4 Conclusions

Advantages of applying the virtual reality in steam generator ALM system have been highlighted in this paper. For most unskilled operators and managers, it is difficult to analyze the information acquired from different components, so a general cruise and 3D indication system for SG is developed. In the environment of virtual reality providing by the developed system, operation, maintenance and installation of the SG can be easily done by all engineering staff. The SG's condition, integral evaluation of SG tube and the latent ageing problems are also shown with the three-dimensional scene. In addition to this, as all information that the staff training needs are displayed in the three-dimension visualization, the system is a very useful tool of training.

Virtual reality is proven to be a cost efficient technique in an ever-tougher commercial world. As the related software technology increase, its ability to simulate the world also increases. VR technology has a wider spectrum from classroom to industrial giant environments.

The presented ALM-VR is now under evaluation in DayaBay Nuclear Power Plant in China, although it is still in the initial stage. Many topics are remained open to study.

References

1. Feiner, S., Thalmann, D.: Virtual Reality [Guest Editor]. *IEEE Computer Graphics and Applications* 20, 24–25 (2000)
2. Aghina, M.A.: Full Scope Simulator of a Nuclear Power Plant Control Room Using 3D Stereo Virtual Reality Technology for Operators Training. In: *Proc. 2007 Int'l Nuclear Atlantic Conf (INAC 2007)*, Associação Brasileira de Energia Nuclear (2007)
3. Hopkins, J.F., Fishwick, P.A.: A Three-dimensional Human Agent Metaphor for Modeling and Simulation. *Proceedings of the IEEE* 89, 131–147 (2001)
4. Guo, J., Li, Z., Chen, T.: Virtual Environment Conception for CBM of Hydro-electric Generating Units. *International Conference on Power System Technology* 3, 1957–1961 (2002)
5. Burdea, G., Patounakis, G., Popescu, V., Weiss, R.: Virtual Reality-based Training for the Diagnosis of Prostate Cancer. *IEEE Trans. on Biomedical Engineering* 46, 1253–1260 (1999)
6. Keyhani, A., Marwali, M.N., Higuera, L.E., Athalye, G., Baumgartner, G.: An Integrated Virtual Learning System for the Development of Motor Drive Systems. *IEEE Trans. on Power Systems* 17, 1–6 (2002)
7. Aghina, M.-A.C., Mol, A.-C.A., Jorge, C., Pereira, C.M.N., Varela, T.F., Cunha, G., Landau, L.: Virtual Control Desks for Nuclear Power Plant Simulation: Improving Operator Training. *IEEE Trans. on Computer Graphics and Applications* 28, 6–9 (2008)
8. Trovato, S.A., Parry, J.O., Burger, J.M.: Slowing the Aging of Nuclear Power Plants. *IEEE Transactions on Spectrum* 32, 32–36 (1995)
9. Guo, J., Li, Z., Chen, Y.: Visualization of a Hydro-electric Generating Unit and Its Applications. *IEEE International Conference on Systems, Man and Cybernetics* 3, 2354–2359 (2003)

10. Banford, H., Fouracre, R.A.: Nuclear Technology and Aging. *IEEE Trans. on Electrical Insulation Magazine* 15, 19–27 (1999)
11. Anandakumaran, K.: Aging and Condition Monitoring Studies of Composite Insulation Cables Used in Nuclear Power Plants. *IEEE Transactions on Dielectrics and Electrical Insulation* 14, 227–237 (2007)
12. Lin, M., Otaduy, M.A., Boulic, R.: Virtual Reality Software and Technology. *IEEE Trans. on Computer Graphics and Applications* 28, 18–19 (2008)
13. Xin, M., Lei, Z., Volkau, I., Weili, Z., Aziz, A., Ang, M.H., Nowinski, W.L.: A Virtual Reality Simulator for Remote Interventional Radiology: Concept and Prototype Design. *IEEE Transactions on Biomedical Engineering* 53, 1696–1700 (2006)
14. Saeedfar, A., Barkeshli, K.: Shape Reconstruction of Three-Dimensional Conducting Curved Plates Using Physical Optics, NURBS Modeling, and Genetic Algorithm. *IEEE Transactions on Antennas and Propagation* 54, 2497–2507 (2006)

Author Index

- Abbas, Ayad R. I-689, II-192
Abbaszadeh Naseri, Mehdi II-1059
Ahmadi, Majid III-337
Alejo, Roberto II-547
Ali, Waleed II-70
Alsharif, Mohammad Reza I-219
Álvarez, Ignacio II-337, III-399
Anbananthen, Kalaiarasi Sonai
Muthu II-520
- Baek, Gyeondong III-1189
Bai, Gang III-416
Bai, Yongjun III-1146
Bao, Haibo I-492
Bao, Huiling I-1161
Bao, Na II-709
Barrón, Ricardo II-977
Beadle, Patch J. I-21, III-530
Beigy, Hamid I-794
Bi, Gexin II-1094
Bian, Yan III-171
Binbin, Huang III-109
Busch, Christoph III-356
- Cai, Lingru III-1122
Cai, Qiufeng I-423
Cai, Yiqiao I-75
Canals, Vincent III-1154
Cao, Buqing II-60
Cao, Chengtao III-1007
Cao, Deguang II-1078
Cao, Feilong III-407
Cao, Jinde I-272, I-492
Cao, Zhongsheng III-380
Challa, Subhash III-449
Chan, Huiling III-512
Chang, De-feng III-152
Chang, Juifang II-794
Chang, Kuang-Chiung I-118
Chao, Kuei-Hsiang I-745
Chaves, Rosa II-337
Chen, Boshan I-601
Chen, Chen II-364
Chen, Chia-Tang I-679
Chen, Chun-Yao II-1116
Chen, Enhong III-49
Chen, Gonggui II-537
Chen, Hao II-172
Chen, Hong II-986
Chen, Huafeng I-512, III-657
Chen, Huaping III-77
Chen, I-Tzu III-512
Chen, Jialiang I-21
Chen, Jiawei II-853, III-457
Chen, Jing III-226
Chen, Li-Fen III-512
Chen, Liujun I-68
Chen, Lu III-1146
Chen, Mianyun III-819
Chen, Qinghua II-853, III-457
Chen, Qiong II-259, III-10
Chen, Rushan III-1054
Chen, Shengbo II-911, III-855
Chen, Shuyue III-638
Chen, Sumei III-162
Chen, Tianping I-323, I-579
Chen, Tzu-Hua III-512
Chen, Weigen I-863
Chen, Xi I-819, III-1222
Chen, Xingang I-863
Chen, Xueguang I-956, I-1144, I-1171,
II-7, II-364, II-969, III-937
Chen, Xueyou II-50
Chen, Yan I-1138
Chen, Yang III-1146
Chen, Yiming III-694
Chen, Yong II-1069
Chen, Yong-Sheng III-512
Chen, Yuehui I-1014
Chen, Yushuo I-1107
Cheng, Cheng III-847
Cheng, Jian I-937
Cheng, Qin II-839
Cheng, Quanxin I-492
Cheng, Weiwei I-707
Cheng, Zunshui II-1197
Choi, Jeoung-Nae II-127
Choi, Kyung-Sik III-257
Chung, TaeChoong II-345

- Cleofas, Laura II-547
 Coit, David W. II-208
 Conde, Guilherme III-1044
 Cruz, Benjamín II-977
 Cui, Guangzhao III-684
 Cui, Lili I-313
 Cui, Shigang III-197
 Cui, Yu III-733
- Dai, Zhicheng III-890
 Dang, Feng III-1212
 Dang, Thi Tra Giang II-1
 Deng, Feiqi I-550
 Deng, Nai-yang II-312
 Deng, Shuo III-899
 Deng, Weihong I-617
 Deng, Xiaolian I-253
 Deng, Zerong II-530
 Deng, Zhidong I-209
 Deng, Zhipo I-877
 de Paúl, Ivan III-1154
 Desai, Sachi III-299
 Ding, Gang II-235
 Ding, Lixin II-60
 Ding, Ming-yue III-1089
 Ding, Yi I-804
 Dong, Fang II-1094
 Dong, Wei I-138
 Dong, Yu III-1106
 Dong, Zhaoyang I-827
 Du, Ji-Xiang II-432, III-136, III-983
 Du, Wei II-382
 Duan, Haibin I-735, III-236
 Duan, Lijuan III-486
 Duan, Miyi III-630
 Duan, Xianzhong II-537
 Duan, Xiyao II-374
- Elek, Istvan I-1053
 Er, Meng Joo II-99
- Faez, Karim III-267
 Fan, Ruiyuan III-188
 Fan, Shuiqing II-43
 Fan, Youping I-813
 Fang, Bin III-1082
 Fang, Fukang II-853, III-457
 Fang, Lei III-88
 Fang, Wei II-225
 Fang, Yanjun I-624
- Fei, Qi I-185, I-1107, I-1115, III-948,
 III-1122, III-1130, III-1222
 Fei, Shumin III-346
 Feng, Guiyu III-630
 Feng, Jian I-440, I-463
 Feng, Lihua I-29
 Feng, Shan I-1072
 Feng, Yi II-859
 Feng, Yong II-684
 Feng, Zhihong II-225
 Francês, Carlos R.L. III-1044
 Fu, Bo III-310
 Fu, Chaojin I-303, I-340
 Fu, Xiaocai II-1
 Fu, Xian I-804
 Fu, Xuezhi I-893, II-875, III-538
 Fu, Xuyun II-235
 Fu, Youming II-1145
 Fung, Chun Che I-175
- Gao, Daming I-844
 Gao, Jiaquan II-500, III-88
 Gao, Jie III-576
 Gao, Meijuan II-555, II-745
 Gao, Meng III-328
 García, Vicente II-547
 Ge, Junfeng I-784
 Ge, Lindong II-737
 Geng, Shuqin I-844, III-772
 Gong, Jingwen II-969
 Gong, Na I-844
 Gong, Qingwu III-874
 Gong, Zhi Chao III-963
 Górriz, Juan M. II-337, III-399
 Gu, Dawu I-60
 Gu, Hong I-836
 Gu, Liangling I-863
 Gu, Suicheng III-466
 Gu, Xueqiang III-226
 Gu, Zhihong II-242, III-1034
 Guan, Liming II-109
 Guo, Baoping III-494
 Guo, Bianjing I-413
 Guo, Chengan II-327
 Guo, Gengqi III-1007
 Guo, Jiang III-1230
 Guo, Jun I-617
 Guo, Libo II-839
 Guo, Ping II-943, II-950
 Guo, Qiyong II-674

- Guo, Quan III-41
 Guo, Sihai I-1072
 Guo, Weizhong III-724
 Guo, Xiufu I-607
 Guo, Xuan III-494
 Guo, Yi-nan I-937
 Guo, Yinbiao II-424
 Guo, Youmin III-973
 Guo, Yue-Fei III-144
 Guo, Zhishan II-480

 Ha, Minghu I-110, I-699
 Habibi, Mehran II-1050
 Han, Honggui III-188
 Han, Qi III-356
 Han, Xinjie II-99
 Han, Yubing III-1054
 He, Guixia II-500
 He, Haibo III-299
 He, Jingjing I-887
 He, Lifeng III-675
 He, Qing II-88
 He, Tong-jun I-164
 He, Xiangnan I-579
 He, Xingui I-670, III-466
 He, Yigang III-714
 He, Zheming I-1033
 He, Zhengping I-1107, I-1115
 Hieu, Duong Ngoc I-52
 Hino, Hideitsu I-84
 Hong, Liu I-956, I-1171, II-364,
 II-753, III-937, III-948
 Hong, Qun I-455
 Hong, Yindie III-371
 Hou, Ligang I-844, III-772
 Hou, TieMin II-364
 Hou, Wen-guang III-1089
 Hu, Biyun I-766
 Hu, Daoyu II-374
 Hu, De-feng III-152
 Hu, Dewen III-630
 Hu, Feng II-839
 Hu, Hong-xian II-591
 Hu, Jiani I-617
 Hu, Jianming III-899
 Hu, Junhao I-512, III-557
 Hu, Liping II-655
 Hu, Mingsheng II-753
 Hu, Rongqiang III-923
 Hu, Ruimin II-1145

 Hu, Sanqing III-605
 Hu, Tao III-494
 Hu, Wan III-1203
 Hu, Wei I-909
 Hu, Weidong I-661
 Hu, Xiaolin III-116
 Hu, Xiaoya III-915, III-956
 Hu, Yabin III-630
 Hu, Yi II-165
 Huang, Hungfu III-1026
 Huang, Jian II-267, III-67
 Huang, Jiangshuai III-67
 Huang, Jinhua I-262, II-218, III-794
 Huang, Lan II-382
 Huang, Qingbao II-1078
 Huang, Qinhuia II-865
 Huang, Wei II-60
 Huang, Wenrong III-1097
 Huang, Xiangzhao I-185
 Huang, Xin-han III-733
 Huang, Zhiwei II-155
 Hüllermeier, Eyke I-707
 Huo, Linsheng I-919
 Huong, Vu Thi Lan I-52

 Itoh, Hidenori III-675

 Jafari, Mohammad II-1013
 Jaiyen, Saichon I-756
 Jamali, Saeed III-267
 Ji, Zhicheng II-1152
 Jia, Jinyuan II-674
 Jia, Zhijuan II-753
 Jian, Jigui I-253, I-395
 Jiang, Bo II-374
 Jiang, Dingguo I-522
 Jiang, Haijun I-413, I-570
 Jiang, Jiang I-1115
 Jiang, Jin III-1230
 Jiang, Jun III-829
 Jiang, Minghui I-560, III-1
 Jiang, Ping III-289
 Jiang, Shuyan III-207
 Jiang, Tao I-887
 Jiang, Weijin II-461
 Jiang, Wenjie I-503
 Jiang, Yi I-819
 Jiang, Zhenhua I-1191
 Jin, Xin I-929
 Jing, Pan-pan III-152

- Jing, Yuanwei I-440
 Jiu, Bo II-304
 Jo, Jun III-1089
 Jou, Chorng-Shyr I-679
 Juan, Liu I-689, II-192

 Kala, Rahul II-821
 Kamel, Mohamed II-276
 Kangshun, Li II-601
 Kazemitabar, Seyed Jalal I-794
 Khorasani, K. III-780
 Khosravy, Mahdi I-219
 Kim, Hyun-Ki I-156
 Kim, Kangkil III-1189
 Kim, Sungshin III-1189
 Kim, Ungmo II-845
 Kim, Yong Soo II-201
 Kim, Yongsu I-313
 Kim, Younghee II-845
 Kong, Li II-986
 Kou, Bingen II-242
 Kou, Guangxing II-80
 Kraipeerapun, Pawalai I-175

 Lai, Xingyu I-635
 Lebbby, Gary III-1112
 Lee, KinHong I-201
 Lee, Suk-Gyu III-257
 Lee, Young-II II-127
 Leong, K.Y. II-520
 Leu, Yih-Guang II-1116, II-1123
 Leung, KwongSak I-201
 Li, Baofeng II-442
 Li, Bing I-919
 Li, Bo II-432, III-983
 Li, Caiwei I-75
 Li, Changhe III-126
 Li, Chaoshun II-155, II-664
 Li, Ching-Ju I-745
 Li, Chuanfeng III-247
 Li, Cuiling III-684
 Li, Di I-1202
 Li, Fenglian III-586
 Li, Fengpan II-155
 Li, Gang II-135, II-904
 Li, Guanjun I-295
 Li, Haobin III-684
 Li, Hongjiao I-60
 Li, Hongnan I-919
 Li, Hongyu II-674
 Li, Hui III-126
 Li, Jie III-299
 Li, Jincheng II-88
 Li, Jing I-60
 Li, Jiuzhong III-1007, III-1137
 Li, Junfang III-1160
 Li, Kenli III-567, III-648
 Li, Lin I-766
 Li, Lishu II-853, III-457
 Li, Liya II-304
 Li, Mingchang I-1
 Li, Ruimin III-1017
 Li, Shujing III-839
 Li, Shusheng III-993
 Li, Ta III-576
 Li, Tai II-1152
 Li, Tan III-829
 Li, Tao II-839
 Li, Wei I-60, III-929, III-1222
 Li, Weiguang I-635
 Li, Wu I-1138
 Li, Xiaobo I-570
 Li, Xiaodong III-346
 Li, Xiaoguang III-684
 Li, Xin III-365
 Li, Xinyu III-1000
 Li, Xiuquan I-209
 Li, Xuechen II-1197
 Li, Yan I-532, II-591, III-152, III-1160
 Li, Yangmin II-1040
 Li, Yanling II-135, II-904
 Li, Ying III-1130
 Li, Yinghai II-664
 Li, Yinhong II-537
 Li, Yuan-xiang II-564
 Li, Zhaoxing I-244
 Li, Zhen II-374
 Li, Zhi III-890
 Li, Zhiyong III-567, III-648
 Li, Zhongxin I-929
 Li, Zhoujun III-694
 Lian, Huicheng III-596
 Liang, Chunlin III-371
 Liang, Hualou II-859, III-365, III-605
 Liang, Jinling I-272
 Liang, Jinming I-405
 Liang, Lishi I-455
 Liang, Shuxiu I-1
 Liang, Yanchun II-382
 Liang, Yue III-217

- Liang, Zhao-hui III-993
 Liao, Hongzhi I-651
 Liao, Jiaping III-310
 Liao, Wudai I-279
 Lin, Jian II-109
 Lin, Jian-You II-1116, II- 1123
 Lin, Xiaofeng II-1078
 Lin, Yu-Jiun III-476
 Lin, Zhigui II-225
 Lin, Zhiqiang III-486
 Liu, Chunming II-398, III-278
 Liu, Derong I-463
 Liu, Desheng II-1138
 Liu, Guangyong III-1122
 Liu, Hesheng III-724
 Liu, Hong III-1203
 Liu, Hongbing II-259, III-10
 Liu, Hongwei II-304, II-655
 Liu, Huaping III-328
 Liu, Jia II-80
 Liu, Jianjun I-661
 Liu, Jianyong II-182
 Liu, Jingjing II-374
 Liu, Jiqing I-262, II-218, III-794
 Liu, Ju III-162
 Liu, June I-1080
 Liu, Junwan III-694
 Liu, Kun-Hong II-424, II-432, III-983
 Liu, Lei III-439
 Liu, Meiqin I-357, I-366
 Liu, Peng III-1230
 Liu, Qing I-1154
 Liu, Qingshan I-272
 Liu, Shan III-207
 Liu, Shubo I-450
 Liu, Suolan III-638
 Liu, Tangbo II-1078
 Liu, Wenju II-928, II-936, III-621
 Liu, Wenxia II-647
 Liu, Wenzhong I-986
 Liu, Xiaobin I-1098
 Liu, Xingcheng II-530
 Liu, Xu I-238
 Liu, Yajin III-1230
 Liu, Yan I-68
 Liu, Yankui II-15, II-25
 Liu, Yansong II-451
 Liu, Yi III-648
 Liu, Yijian I-624
 Liu, Yong III-126
 Liu, Yu II-611
 Liu, Zhiqiang II-25
 Liu, Zhong I-893, II-875, III-217,
 III-538, III-1063
 Liu, Zhuo III-915, III-956
 Long, Aifang I-194
 Long, Fei II-1023, II-1032
 Long, Hao III-1071
 Long, Xingming I-104
 López, Miriam II-337, III-399
 Lou, Suhua II-611
 Lu, Bao-Liang II-784
 Lu, Huapu III-1017
 Lu, Jian I-185
 Lu, Junan II-1130
 Lu, Li II-480
 Lu, WenBing II-694
 Lu, Wenlian I-323, I-579
 Lu, Youlin II-664, III-30
 Lu, Zhongkui II-1069
 Luo, Fei III-1197
 Luo, Li III-310
 Luo, Siwei I-728
 Luo, Yasong I-893, III-538, III-1063
 Luo, Youxin I-1033
 Luo, Yupin I-784, III-390
 Luo, Zhenguo I-383
 Lursinsap, Chidchanok I-756
 Ma, Chao II-784
 Ma, Dan I-244
 Ma, Jinwen II-959
 Ma, Liying III-780
 Ma, Shoufeng III-1082
 Ma, Xiaohong II-859, III-365
 Ma, Zhenghua III-638
 Makaremi, Iman III-337
 Malek, Alaeddin III-98
 Malekjamshidi, Zahra II-1013
 Man, Hong III-299
 Mao, Cheng-xiong II-591, III-152,
 III-1160
 Mao, Yuming III-909
 Mao, Zijun III-937, III-948
 Meng, Jinsong III-207
 Meng, Ke I-827
 Meng, Qingfang I-1014
 Meng, Song II-99
 Meng, Xianyao II-99
 Miao, Jun III-486

- Miao, Xiao-yang III-152
 Minghui, Wang II-1189
 Miyajima, Hiromi II-118, II-886
 Moghbelli, Hassan I-852
 Moran, Bill III-449
 Morro, Antoni III-1154
 Murata, Noboru I-84
- Nagamine, Shinya II-118
 Nakamura, Tsuyoshi III-675
 Nakkrasae, Sathit I-175
 Ngamwitthayanon, Nawa II-208
 Nguyen, Minh Nhut II-1
 Ning, Bo I-870
 Ning, Di I-194
 Ning, Xiaoling III-1063
 Ninh, Sai Thi Hien I-52
 Niu, Dongxiao II-242, III-1034
 Niu, Guang-dong I-937
 Niu, Xiamu III-356
 Niu, Xinxin III-318
- Oh, Sung-Kwun I-156, II-127
 Ou, Yangmin I-185
 Ouyang, Min III-937
 OuYang, Ming III-948
 Ouyang, Weimin II-865
 Ozlati Moghadam, Mostafa III-267
- Pan, Chen III-407
 Pang, Chuanjun III-839
 Pang, Hali I-1098
 Park, Dong-Chul I-52, I-967
 Park, Ho-Sung I-156
 Parkkinen, Jussi II-674
 Peng, Hui I-870
 Peng, Lingxi III-371
 Peng, Pengfei II-875, III-538
 Peng, Shouye II-928, II-936
 Peng, Wen II-647
 Peng, Xiaohong I-844
 Peng, Xufu I-804
 Peng, Zhaoxia I-909
 Peng, Zhenrui III-973
 Pengcheng, Li II-601
 Phimoltares, Suphakant I-756
 Ping, Huang II-601
 Puntonet, Carlos G. III-399
- Qasem, Sultan Noman III-19
 Qi, Chuanda II-904
- Qi, Hang II-763
 Qi, Xiaowei I-774
 Qi, Xinbo III-684
 Qian, Hai I-813
 Qian, Jian-sheng I-937
 Qiao, Junfei III-188
 Qiao, Xing I-244
 Qiao, Yuanhua III-486
 Qin, Hui II-664, III-30
 Qin, Ling III-923
 Qin, Rui II-25
 Qin, Taigui I-1024
 Qiu, Meikang I-357, I-366
- Rahideh, Akbar I-852
 Ramírez, Javier II-337, III-399
 Ranjan, Anand II-821
 Rao, Congjun I-1080, I-1090,
 I-1131, I-1161
 Rao, Hao II-490
 Rego, Liviane P. III-1044
 Ren, Fujun I-94
 Ren, Guang I-450, I-774
 Ren, Min III-226
 Rocha, Cláudio A. III-1044
 Rosselló, Josep L. III-1154
 Ruan, Dianxu II-251
 Ruan, Gongqin I-36
 Ruan, Qian II-581
 Ruan, Xiaogang III-188
 Ruan, Xin-bo II-591, III-152, III-1160
 Ruxpakawong, Phongthep I-229
- Safavi, Ali A I-852
 Salas-Gonzalez, Diego II-337, III-399
 Sang, Haifeng II-831
 Santana, Ádamo L. de III-1044
 Segovia, Fermín II-337, III-399
 Shamsuddin, Siti Mariyam II-70, III-19
 Shang, Fengjun III-809
 Shang, Peng II-510, III-864
 Shao, Fengjing III-839
 Shen, Hui III-171
 Shen, Lei I-766
 Shen, Lincheng III-226
 Shen, Minfen I-21
 Shen, Tsurng-Jehng I-679
 Shen, Xianfeng III-1097
 Shen, Yanjun I-347
 Shen, Yanxia II-1152
 Shen, Yi III-1

- Shi, Bao I-472
 Shi, Bin III-41
 Shi, Daming II-1
 Shi, Muyao I-503
 Shi, Qingjun II-1138
 Shi, Shiyong III-909
 Shi, Shuo III-1007, III-1137
 Shi, Weiya III-144
 Shi, Xi III-874
 Shi, Zhenghao III-675
 Shi, Zhengping I-164
 Shi, Zhongzhi II-88
 Shigei, Noritaka II-118, II-886
 Shu, Feng III-1054
 Shukla, Anupam II-821
 Silva, Marcelino S. da III-1044
 Sitiol, Augustina II-520
 Song, Bifeng II-442
 Song, Bong-Keun III-257
 Song, Guojie I-670
 Song, Haigang I-956, I-1144, I-1171,
 II-7, II-364, II-753, II-969
 Song, Hu III-1106
 Song, Jiekun II-43
 Song, Jiepeng II-43
 Song, Jinze II-398
 Song, Qiankun I-405, I-482, I-542
 Song, Shujie III-1071
 Song, Y.D. III-1112
 Sossa, Humberto II-977, III-520
 Stead, Matt III-605
 Su, Gang II-510, III-829, III-864
 Su, Hongsheng II-172
 Su, Jianmin II-442
 Su, Lei I-651
 Su, Zhewen III-380
 Sui, Yan III-557
 Sun, Changyin II-727
 Sun, Fuchun II-480, III-328
 Sun, Haishun I-1154
 Sun, Lisha III-530
 Sun, Qiaomei I-774
 Sun, Qun II-322
 Sun, Rencheng III-839
 Sun, Shiliang II-802, II-996
 Sun, Xichao III-1212
 Sun, Yinjie III-59
 Sun, Yuehui III-429
 Sun, Zengqi II-1180
 Sun, Zhaochen I-1
 Sun, Zheng II-251
 Sun, Zigang III-819
 Suo, Hongbin II-639
 Taghvae, Sajjad III-267
 Tai, Shenchuan III-1026
 Tan, Jian II-702
 Tan, Li II-510, III-864
 Tan, Manchun I-433
 Tan, Ning I-11
 Tan, Ying III-466
 Tang, Jiahua I-651
 Tang, Jian III-546
 Tang, Pan I-1181
 Tang, Qiang III-890
 Tang, Qin III-126
 Tang, Yun II-928
 Tao, Jing III-621
 Tassing, Remi III-663
 Teng, Zhidong I-413
 Thammano, Arit I-229
 Thuy, Nguyen Thi Thanh II-345
 Tian, Gang II-1145
 Tian, Jingwen II-555, II-745
 Tian, Liguo III-197
 Tian, Ying-jie II-312
 Tiwari, Ritu II-821
 Tong, Xiaoqin III-310
 Tsai, Cheng-Fa III-476
 Tu, Lilan II-1005
 Utiyama, Masao II-784
 Valdes, Arturo III-780
 Valdovinos, Rosa Maria II-547
 Vázquez, Roberto A. III-520
 Vien, Ngo Anh II-345
 Viet, Nguyen Hoang II-345
 Vo, Nhat III-449
 Wan, Feng II-354
 Wan, Jiafu I-1202
 Wan, Lei I-986
 Wang, Bin II-737
 Wang, Bingwen III-890, III-915, III-956
 Wang, Boyu II-354
 Wang, Chao I-110, I-699
 Wang, Cheng I-1062, I-1090
 Wang, Chengshan III-1171
 Wang, Chuncai II-382
 Wang, Desheng III-663

- Wang, Dongyun I-286
 Wang, Fei II-709
 Wang, Fushan II-1180
 Wang, Guoli II-15, II-25, II-80
 Wang, Guoyou II-921
 Wang, Guozheng I-383
 Wang, Haipeng II-639
 Wang, Heyong II-621
 Wang, Honggang I-827
 Wang, Hongwei I-819, I-836, I-946,
 I-1041, I-1123, I-1181, III-178
 Wang, Jiahai I-75
 Wang, Jian I-946, I-1041
 Wang, Jianfeng I-253, I-395
 Wang, Jianqin I-542
 Wang, Jianyong III-657
 Wang, Jiao I-728
 Wang, Jinfeng I-201, III-847
 Wang, Jing I-94, I-607, III-371
 Wang, Jinhui I-844, III-772
 Wang, Juexin II-382
 Wang, Jun I-766
 Wang, Kai II-591
 Wang, Kuanquan III-439
 Wang, Laisheng II-322, II-631
 Wang, Lei II-287, II-1160, II-1165
 Wang, Lijuan III-1212
 Wang, Lili I-323
 Wang, Lixin II-859
 Wang, Long I-94
 Wang, Luyao II-374
 Wang, Meng-Huei I-745
 Wang, Min III-733
 Wang, Mingchang II-911
 Wang, Nan III-226
 Wang, Ning II-99
 Wang, Qian II-391
 Wang, Qing I-1131, I-1161
 Wang, Qiwan II-572
 Wang, Rubin I-138
 Wang, Rulong II-451
 Wang, Shan I-1098
 Wang, Shouxiang III-1171
 Wang, Shuqing II-145, III-762
 Wang, Ti-Biao II-1165
 Wang, Tuo I-503
 Wang, Wei II-225
 Wang, Wentao II-921
 Wang, Xianjia II-702
 Wang, Xiao-Guo III-423
 Wang, Xiaoping II-581
 Wang, Xingjun III-1097
 Wang, Xiuhua II-172
 Wang, Xuan II-564
 Wang, Yan I-286, II-382
 Wang, Yanfeng III-684
 Wang, Yanran III-236
 Wang, Yanwu II-839
 Wang, Yifan II-727
 Wang, Yin III-899
 Wang, Ying III-30
 Wang, Yong I-60
 Wang, Yongji II-267, III-67,
 III-207, III-247
 Wang, Youyi I-827
 Wang, Yu I-1171, II-969, III-546
 Wang, Yuanmei II-839
 Wang, Yuanzhen III-380
 Wang, Zhanshan I-440, I-463
 Wang, Zhe I-819
 Wang, Zhifang III-356
 Wang, Zhiwu I-1144
 Wang, Zhongsheng I-279, I-340
 Wang, Zhongyuan II-1145
 Wang, Zijun II-911
 Wattanapongsakorn, Naruemon II-208
 Wei, Cheng III-744
 Wei, Lin II-1189
 Wei, Xingxing III-236
 Wei, Yongchang I-1123
 Wei, Yongjun I-1098
 Weibing, Liu III-109
 Wen, Bin II-60
 Wen, Guoguang I-909
 Wen, Jinyu I-1154, III-1179
 Wen, Lin I-893
 Weng, Guoqing II-165
 Weng, Liguo III-1112
 Woo, Dong-Min I-52, I-967
 Worrell, Gregory A. III-605
 Wu, Ailong I-303
 Wu, Charles Q. III-502
 Wu, Chong I-976
 Wu, Chunmei II-470
 Wu, Chunxue I-472
 Wu, Dongqing I-455
 Wu, Hao III-247
 Wu, Jiansheng II-470, III-49
 Wu, Jie II-1087
 Wu, Jing I-110

- Wu, Jiutao I-870
 Wu, Jun II-267
 Wu, Kaigui II-684
 Wu, Qing Ming III-963
 Wu, Shunjun II-304, II-655
 Wu, Tianshu I-670
 Wu, Wuchen I-844, III-772
 Wu, Yanyan I-149
 Wu, Yaowu II-611
 Wu, Zhongfu II-684

 Xi, Shijia II-480
 Xia, Bin III-612
 Xia, Shengping I-661
 Xia, Yongbo III-557
 Xia, Youshen I-877, II-276
 Xiang, Yang II-35
 Xianjia, Wang III-109
 Xiao, Degui III-567, III-648
 Xiao, Feng III-1089
 Xiao, Huabiao II-943
 Xiao, Huimin I-375
 Xiao, Jianhua III-704
 Xiao, Li II-839, III-819
 Xiao, Longteng III-171
 Xiao, Zhihuai II-145
 Xiao, Zhitao II-225
 Xiaoxu, Yan III-744
 Xie, Kunqing I-670
 Xie, Qingguo II-374
 Xie, Shutong II-424
 Xie, Yinghui II-1180
 Xie, Yongle III-207
 Xie, Zhipeng II-773
 Xin, Youming II-1197
 Xing, Jianmin II-1197
 Xing, Jun II-875
 Xing, Lining II-490
 Xing, Lixin II-911
 Xing, Tingyan I-503
 Xing, Zhihui I-735
 Xiong, Guoliang III-724
 Xiong, Shengwu II-259, III-10
 Xiong, Zhaogang I-601
 Xu, Chao III-1097
 Xu, Chunfang I-735
 Xu, Guiyun II-251
 Xu, Jian-Hao II-1165
 Xu, Jie I-717
 Xu, Jin I-295

 Xu, Jing III-596
 Xu, Qi II-267
 Xu, Qingsong II-1040
 Xu, Qingyang II-99
 Xu, Wei I-601, II-182
 Xu, Xiaodong I-238
 Xu, Xin II-398, III-278
 Xu, Xiuli II-950
 Xu, Xuelian III-197
 Xu, Yitian II-322, II-631
 Xu, Yong II-424, II-1160
 Xu, Yuge III-1197
 Xu, Yuhui II-461
 Xu, Zhiming III-546
 Xu, Zhiru II-1138
 Xue, Fuqiang II-737
 Xue, Lin III-162
 Xue, Liqin III-762
 Xue, Zhibin II-1105

 Yamashita, Katsumi I-219
 Yan, Chunyan I-635
 Yan, Hehua I-1202
 Yan, Liexiang III-41
 Yan, Peng I-395
 Yan, Shujing I-699
 Yan, Xunshi III-390
 Yan, Yonghong II-639, III-576
 Yang, Bin I-1014
 Yang, Chao II-591
 Yang, Chunyan II-911
 Yang, Fengjian I-455
 Yang, Genghuang III-197
 Yang, Hai III-909
 Yang, Haicong III-801
 Yang, Huimin III-1179
 Yang, Jianfu I-455
 Yang, Jianhua III-1160
 Yang, Jianxi I-482
 Yang, Lei III-567, III-648
 Yang, Liming II-322, II-631
 Yang, Ou III-494
 Yang, Qing III-1203
 Yang, Qingshan II-327
 Yang, Wenjun III-956
 Yang, Xuezhao I-279
 Yang, Yan II-959
 Yang, Yi I-863
 Yang, Yiwen I-11
 Yang, Yixian III-318

- Yang, Yong II-287
 Yang, Yongli III-663
 Yang, Yongqing II-1171
 Yang, Zhi III-1122, III-1130
 Yao, Hongshan II-408, II-416
 Yao, Minghui I-976
 Yao, Ping I-976, I-993
 Yao, Yong-feng II-591
 Yao, Zhang II-601
 Yashtini, Maryam III-98
 Yatsuki, Shuji II-886
 Yazdizadeh, Alireza II-1050, II-1059
 Ye, Bangyan I-635
 Ye, Chunxiao II-684
 Ye, Hongtao III-1197
 Ye, Lin I-589, III-929
 Yeh, Ming-Feng I-118
 Yi, Daqing III-289
 Yi, Gang II-451
 Yi, Jiangan III-755
 Yin, An III-915
 Yin, Jian I-75
 Yin, Jianchuan II-1094
 Yin, X.H. III-1112
 Yin, Yean II-694
 Yongquan, Yu II-1189
 Youh, Meng-Jey I-679
 Yu, Jianjiang I-423
 Yu, Jun III-530
 Yu, Lin I-46
 Yu, Shiwei I-607
 Yu, Shuanghe III-178
 Yu, Simin II-717, II-810
 Yu, Weiyu II-717
 Yu, Wenxian I-661
 Yu, Ying II-276
 Yu, Yongguang I-909
 Yu, Zhengtao I-651
 Yu, Zhiding II-717, II-810
 Yuan, Jingling I-887
 Yuan, Qiping III-278
 Yuan, Tiantian II-895
 Yuan, Weiqi II-831
 Yuan, Xiaohui II-145, III-762
 Yuchi, Ming III-1089

 Zeng, An I-68
 Zeng, Bin I-1033
 Zeng, Jianchao II-1105
 Zeng, Jianyou III-126

 Zeng, Maimai I-1062
 Zeng, Peng III-755
 Zeng, Wei I-1107, I-1115, I-1181,
 III-1122, III-1130
 Zeng, Yanshan I-455
 Zhai, Chuan-Min III-136
 Zhan, Xisheng II-1087
 Zhan, Yunjun I-149
 Zhang, Bo III-116
 Zhang, Bu-han II-591, III-152, III-1160
 Zhang, Chaolong I-455
 Zhang, Daming II-1069
 Zhang, David III-439
 Zhang, Dexian II-709
 Zhang, Di III-429
 Zhang, Faming I-333
 Zhang, Fan II-555
 Zhang, Gan-nian II-564
 Zhang, Gaolei III-1171
 Zhang, Guangyu I-1
 Zhang, Guojun III-310
 Zhang, Hua II-936
 Zhang, Huaguang I-313, I-463
 Zhang, Hui III-829
 Zhang, Jianhong III-318
 Zhang, Jianwei I-209
 Zhang, Jie I-589, III-929
 Zhang, Jun II-432, III-983
 Zhang, Ke II-694
 Zhang, Lingmi II-296
 Zhang, Linguo I-347
 Zhang, Linlan II-7
 Zhang, Liqing II-763
 Zhang, Long III-724
 Zhang, Mingwang III-77
 Zhang, Nan I-503, II-611
 Zhang, Pengcheng II-398, III-278
 Zhang, Ping I-1202
 Zhang, Qi III-923
 Zhang, Qiang III-963
 Zhang, Qianhong I-383
 Zhang, Qihua I-956, II-969
 Zhang, Qunjiao II-1130
 Zhang, Rui I-440
 Zhang, Senlin I-357, I-366
 Zhang, Sheng II-1023, II-1032
 Zhang, Shishuang II-530
 Zhang, Tao III-346
 Zhang, Tianxu II-921
 Zhang, Wang I-844

- Zhang, Wenle I-1002
 Zhang, Xiang II-639
 Zhang, Xianhe II-1087
 Zhang, Xiankun I-110, I-699
 Zhang, Xiaoguang II-251
 Zhang, Xiaomei I-1138
 Zhang, Xin I-313
 Zhang, Xingcai I-29
 Zhang, Xueying III-586
 Zhang, XuncaI III-684
 Zhang, Yanhui I-1144
 Zhang, Yi III-899
 Zhang, Yong I-347
 Zhang, Yongchuan II-155, III-30
 Zhang, Youbing II-165
 Zhang, Yu II-43
 Zhang, Yuan I-450
 Zhang, Yuanyuan III-874, III-909
 Zhang, Yunong I-11, I-36, I-75
 Zhang, Yunyun II-242, III-1034
 Zhang, Zaixu II-43
 Zhang, Zhen I-589
 Zhang, Zheng II-895, III-416
 Zhang, Zhi II-374
 Zhang, Zhijia II-831
 Zhang, Zhijing I-929
 Zhang, Zhikang I-138
 Zhang, Zhi Qiang III-963
 Zhang, Zhongcheng I-643,
 I-1080, I-1090
 Zhang, Zipeng II-145, III-762
 Zhao, Birong I-550
 Zhao, Ge III-1171
 Zhao, Haina II-727
 Zhao, Jianye I-870
 Zhao, Kun II-312
 Zhao, Li III-171, III-197
 Zhao, Liping III-567
 Zhao, Qi III-630
 Zhao, Qingwei II-639, III-576
 Zhao, Xinquan I-128
 Zhao, Yanhong II-1171
 Zhao, Yong I-1072, I-1131,
 I-1161, III-1146
 Zhao, Yongqing I-560, III-1
 Zhao, Yuqin III-77
 Zhao, Zheng II-895, III-416
 Zhen, Ling II-631
 Zheng, Mingfa II-15, II-80
 Zheng, Xiaoming III-390
 Zheng, Ying I-589, III-929
 Zhi, Jun II-182
 Zhi, Limin II-182
 Zhong, Jiang II-684
 Zhong, Jingjing I-728
 Zhong, Luo I-887
 Zhong, Shisheng II-235
 Zhou, Aijun I-450
 Zhou, Bin I-1
 Zhou, Chunguang II-382
 Zhou, Hongtao I-1107, I-1115,
 III-1122, III-1130
 Zhou, Hui II-267
 Zhou, Jian-Lan III-882
 Zhou, Jianting I-482
 Zhou, Jianzhong II-155, II-664, III-30
 Zhou, Jing I-104
 Zhou, Jingguo II-1138
 Zhou, Kaibo I-1072
 Zhou, Lin III-1230
 Zhou, Ming I-986
 Zhou, Renlai III-638
 Zhou, Shiru II-745
 Zhou, Yiming I-766
 Zhousuo, Zhang III-744
 Zhu, Desen III-819
 Zhu, Guangxi II-510, III-663,
 III-829, III-864
 Zhu, Jin III-289
 Zhu, Jingguo III-546
 Zhu, Kejun I-607
 Zhu, Liqiang I-901
 Zhu, Luo III-310
 Zhu, Mei I-94
 Zhu, Wenji III-714
 Zhu, Yue II-296
 Zou, Bin I-717
 Zou, Huijun III-724
 Zou, Ling III-638
 Zou, Ruobing II-717, II-810
 Zuo, Fuchang I-929
 Zuo, Lei I-844, III-772
 Zuo, Wangmeng III-439