

Hepu Deng
Duoqian Miao
Fu Lee Wang
Jingsheng Lei (Eds.)

Communications in Computer and Information Science

237

Emerging Research in Artificial Intelligence and Computational Intelligence

International Conference, AICI 2011
Taiyuan, China, September 2011
Proceedings

Communications
in Computer and Information Science

237

Hepu Deng Duoqian Miao Fu Lee Wang
Jingsheng Lei (Eds.)

Emerging Research in Artificial Intelligence and Computational Intelligence

International Conference, AICI 2011
Taiyuan, China, September 23-25, 2011
Proceedings

Volume Editors

Hepu Deng
RMIT University
Melbourne, VIC, Australia
E-mail: hepu.deng@rmit.edu.au

Duoqian Miao
Tongji University
Shanghai, China
E-mail: miaoduoqian@163.com

Fu Lee Wang
Caritas Institute of Higher Education
Hong Kong, China
E-mail: pwang@cihe.edu.hk

Jingsheng Lei
Shanghai University of Electric Power
Shanghai, China
E-mail: jshlei@126.com

ISSN 1865-0929

e-ISSN 1865-0937

ISBN 978-3-642-24281-6

e-ISBN 978-3-642-24282-3

DOI 10.1007/978-3-642-24282-3

Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011936652

CR Subject Classification (1998): I.2, H.4, I.4-5, F.1, H.3

© Springer-Verlag Berlin Heidelberg 2011

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.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The 2011 International Conference on Artificial Intelligence and Computational Intelligence (AICI 2011) was held during September 23–25, 2011 in Taiyuan, China. AICI 2011 received 1,073 submissions from 20 countries and regions. After rigorous reviews, 265 high-quality papers were selected for publication in the AICI 2011 proceedings. The acceptance rate was 24%.

The aim of AICI 2011 was to bring together researchers working in many different areas of artificial intelligence and computational intelligence to foster the exchange of new ideas and promote international collaborations. In addition to the large number of submitted papers and invited sessions, there were several internationally well-known keynote speakers.

On behalf of the Organizing Committee, we thank Taiyuan University of Technology for its sponsorship and logistics support. We also thank the members of the Organizing Committee and the Program Committee for their hard work. We are very grateful to the keynote speakers, session chairs, reviewers, and student helpers. Last but not least, we thank all the authors and participants for their great contributions that made this conference possible.

September 2011

Hepu Deng
Duoqian Miao
Fu Lee Wang
Jingsheng Lei

Organization

Organizing Committee

General Co-chairs

Wendong Zhang
Qing Li

Taiyuan University of Technology, China
City University of Hong Kong, Hong Kong

Program Committee Co-chairs

Hepu Deng
Duoqian Miao

RMIT University, Australia
Tongji University, China

Steering Committee Chair

Jingsheng Lei

Shanghai University of Electric Power, China

Local Arrangements Co-chairs

Fu Duan
Dengao Li

Taiyuan University of Technology, China
Taiyuan University of Technology, China

Proceedings Co-chairs

Fu Lee Wang
Ting Jin

Caritas Institute of Higher Education,
Hong Kong
Fudan University, China

Sponsorship Chair

Zhiyu Zhou

Zhejiang Sci-Tech University, China

Program Committee

Adi Prananto
Adil Bagirov
Ahmad Abareshi
Alemayehu Molla
Andrew Stranier
Andy Song
An-Feng Liu

Swinburne University of Technology, Australia
University of Ballarat, Australia
RMIT University, Australia
RMIT University, Australia
University of Ballarat, Australia
RMIT University, Australia
Central South University, China

VIII Organization

Arthur Tatnall	Victoria University, Australia
Bae Hyeon	Pusan National University, Korea
Baoding Liu	Tsinghua University, China
Carmine Sellitto	Victoria University, Australia
Caroline Chan	Deakin University, Australia
CheolPark Soon	Chonbuk National University, Korea
Chowdhury Morshed	Deakin University, Australia
Chung-Hsing Yeh	Monash University, Australia
Chunqiao Tao	South China University, China
Costa Marly	Federal University of Amazonas, Brazil
Craig Parker	Deakin University, Australia
Daowen Qiu	Zhong Shan University, China
Dat Tran	University of Canberra, Australia
Dengsheng Zhang	Monash University, Australia
Edmonds Lau	Swinburne University of Technology, Australia
Elsbeth McKay	RMIT University, Australia
Eng Chew	University of Technology Sydney, Australia
Feilong Cao	China Jiliang University, China
Ferry Jie	RMIT University, Australia
Furutani Hiroshi	University of Miyazaki, Japan
Gour Karmakar	Monash University, Australia
Guojun Lu	Monash University, Australia
Heping Pan	University of Ballarat, Australia
Hossein Zadeh	RMIT University, Australia
Ian Sadler	Victoria University, Australia
Irene Zhang	Victoria University, Australia
Jamie Mustard	Deakin University, Australia
Jeff Ang Charles	Darwin University, Australia
Jennie Carroll	RMIT University, Australia
Jenny Zhang	RMIT University, Australia
Jian Zhou T.	Tsinghua University, China
Jingqiang Wang	South China University, China
Jinjun Chen	Swinburne University of Technology, Australia
Joarder Kamruzzaman	Monash University, Australia
Kaile Su	Beijing University, China
Kankana Chakrabaty	University of New England, Australia
Konrad Peszynski	RMIT University, Australia
Kuoming Lin	Kainan University, Taiwan
Lemai Nguyen	Deakin University, Australia
Leslie Young	RMIT University, Australia
Liping Ma	University of Ballarat, Australia
Luba Torline	Deakin University, Australia
Maple Carsten	University of Bedfordshire, UK
Maria Indrawan	Monash University, Australia
Peter Shackleton	Victoria University, Australia
Philip Branch	Swinburne University of Technology, Australia

Pradip Sarkar	RMIT University, Australia
Qiang Li	University of Calgary, Canada
Ravi Mayasandra	RMIT University, Australia
Richard Dazeley	University of Ballarat, Australia
Sanming Zhou	University of Melbourne, Australia
Santoso Wibowo	RMIT University, Australia
Schetinin Vitaly	University of Bedfordshire, UK
Shengxiang Yang	University of Leicester, UK
ShyhWei Teng	Monash University, Australia
Siddhi Pittayachawan	RMIT University, Australia
Stephen Burgess	Victoria University, Australia
Sungshin Kim	Pusan National University, Korea
Syed Nasirin	Brunel University, UK
Tae-Ryong Jeon	Pusan National University, Korea
Tayyab Maqsood R.	MIT University, Australia
Tony Zhang	Qingdao Univesity, China
Vanessa Cooper	RMIT University, Australia
Wei Lai	Swinburne University of Technology, Australia
Wei Peng	RMIT University, Australia
Weijian Zhao	China Jiliang University, China
Xiaodong Li	RMIT University, Australia
Xiaohui Zhao	Swinburne University of Technology, Australia
Yan-Gang Zhao	Nagoya Institute of Technology, Japan
Yang-Cheng Lin	National Dong Hwa University, Taiwan
Yi-Hua Fan	Chung Yuan Christian University Taiwan, Taiwan
Yuan Miao	Victoria University, Australia
Yubin Zhong	Guangzhou University, China
Yubo Yuan	China Jiliang University, China
Yuefeng Li	Queensland University of Technology, Australia
Zhaohao Sun	University of Ballarat, Australia
Zhichun Wang	Tianjin University, China

Table of Contents

Applications of Artificial Intelligence

Optimization Model of Rotation Irrigation Channel Distribution with GA and FS	1
<i>Weizeng Gao, Zhou Yu, and Guoyi Miao</i>	
Search Engine Based Maintainability Evaluation Measure of Web Site	9
<i>Xiao-dong Wang and Jie Cui</i>	
Research on Modeling Plant Growth in Greenhouse Based on Virtual Plant Technology	17
<i>Tang Wei-dong, Li Jin-zhong, and Hu Xue-hua</i>	
Credit Assessment with Random Forests	24
<i>Lei Shi, Yi Liu, and Xinming Ma</i>	
A Study of Dynamic Decoupling Fuzzy Control Systems by Means of Generalized Inverse Matrix	29
<i>Qing-Min Fan</i>	

Applications of Computational Intelligence

Fractal Traffic Analysis and Applications in Industrial Control Ethernet Network	34
<i>Sen-xin Zhou, Jiang-hong Han, and Hao Tang</i>	
Routing on a Spherical Surface Using Hybrid PSO	43
<i>Shoubao Su, Shuhao Yu, Yan Ma, Yang Yang, and Huali Xu</i>	
An Algorithm of Program Comprehension and Visual Representation for Object-Oriented Program	52
<i>Hui Gu and Daomiao Lin</i>	
The Application of Simulated Algorithm Based on .NET in NP	60
<i>Ailian Wang and Yueying Duan</i>	
Coevolutionary Optimization Algorithm: With Ecological Competition Model	68
<i>Jianguo Liu and Weiping Wu</i>	
Secure Secret-Key Management of Kerberos Service	76
<i>Lai-Cheng Cao</i>	

Application of Shuangchao Hydrological Model in Shanxi Semiarid and Semi-humid Area 84
Yutao Cao and Xihuan Sun

The Design of Withstand Test Simulation System Based on Resonant Principle 92
Yonggang Li, Xia Ping, Chen Zhao, and Shuting Wan

Improving Depth-First Search Algorithm of VLSI Wire Routing with Pruning and Iterative Deepening 100
Xinguo Deng, Yangguang Yao, and Jiarui Chen

The Forecast of Our Private Car Ownership with GM(1,1) Model 108
Yi Zhang

The Rationality Properties of Three Preference-Based Fuzzy Choice Functions 113
Xue Na, Yixuan Ma, and Yonghua Hao

Automated Problem Solving

The Application of Internet of Things Technologies in Transmission Link of Smart Grid 120
Yu Jun and Zhang Xueying

Assess of the Flicker Caused by Electric Arc Furnace Using Simulation Method 128
Wang Yufei, Hua Xue, and Yu Xiao

Research to Prevent the Anti-charging Incident of the Voltage Transformer in the Substation 135
Hui-Ying Chen, Zhi-Peng Wang, and Mu-Qin Tian

Research on Online Monitoring Method for Longitudinal Rip of Steel-Core Belt 141
Qiao Tie-Zhu, Wang Fu-Qiang, and Lu Xiao-Yu

Brain Models/Cognitive Science

Establishment and Analysis of Information Processing Models on Intelligent Overtaking Behavior 147
Xin Gao, Gang Dong, and Li Gao

Inhibition of Return and Stroop Effect in Spatial Attention 155
Jia Fu

Data Mining and Knowledge Discovering

Two Dimensions Data Slicing Algorithm, a New Approach in Mining Rules of Literature in Traditional Chinese Medicine	161
<i>Guang Zheng, Hongtao Guo, Yuming Guo, Xiaojuan He, Zhongxian Li, and Aiping Lu</i>	
Frequent Closed Pattern Mining Algorithm Based on COFI-Tree	175
<i>Jihai Xiao, Xiaohong Cui, and Junjie Chen</i>	
Improved Algorithm for Mining N-most Interesting Itemsets	183
<i>Xiaohong Cui, Jihai Xiao, Junjie Chen, and Lijun Sang</i>	
An Improved KFCM Algorithm Based on Artificial Bee Colony	190
<i>Xiaoqiang Zhao and Shouming Zhang</i>	
Improved Clustering Algorithm Based on Local Agglomerative Characteristics	199
<i>Xi-xian Niu and Yan-ping Cui</i>	
Cluster Analysis Based on GAPSO Evolutionary Algorithm	207
<i>Tieqiao Huo, Junxi Zhang, and Xiaojun Wu</i>	
A Model of a GEP-Based Text Clustering on Counter Propagation Networks	214
<i>Jin'guang Luo, Chang'an Yuan, and Jinkun Luo</i>	
The Application of Data Mining Technology in Analysis of the Relation between Majors, Industries and Positions	222
<i>Xiaoguo Wang and Lin Sun</i>	
Research on Partner Selection Issue of Manufacturing Enterprise under Supply Chain Environment	230
<i>Hong Zhang and Feipeng Guo</i>	
Formal Concept Analysis Based on Rough Set Theory and a Construction Algorithm of Rough Concept Lattice	239
<i>Haifeng Yang</i>	

Expert and Decision Support Systems

Command-to-Submarine Decision Model Research of Formation's Cooperative Anti-Submarine Warfare	245
<i>Jinping Wu, Yi Hu, and Rugang Song</i>	
A Context-Aware Model Based on Mobile Agent	253
<i>Ying Wang and Xinguang Peng</i>	
A Mode of Storm Flood Forecasting DSS Establish Ion	261
<i>Yaoping Yan and Yutao Cao</i>	

Research on Cut Order Planning for Apparel Mass Customization 267
Liu Yan-mei, Yan Shao-cong, and Zhang Shu-ting

Fuzzy Logic and Soft Computing

The Semantics of Dynamic Fuzzy Logic Programming Language 272
Xiaofang Zhao

Trend Prediction of Oil Temperature for Wind Turbine Gearbox Based on Grey Theory 280
Wang Rui and Li Gang

The Semantics of Logic System $lp(X)$ 286
Hua Li

Design of Field Integrative Irrigation Control System Based on Fuzzy Control and PLC 295
Xiumei Jia, Lili Yao, and Yingmei Zhang

Intelligent Agents and Systems

Research on the Information Integration Model Based on Multi-agent . . . 302
Zhang Yingjie, Ma Quanzhong, Wen Liangjun, Li Guodong, and Hu Pengfei

Hardware Design of T/Q/IQ/IT for Intra 1616 in H.264/AVC 310
Meihua Gu and Jinyan Hao

Construction of Multi-Agent System for Decision Support of Online Shopping 318
Jiao Li and Yuqiang Feng

Intelligent Control

Robot Real-Time Motion Planning and Collision Avoidance in Dynamically Changing Environments 325
Zhang Jin-xue

Design of the On-Line Detection System for the Surface of the Hot Heavy Rail 335
Renxi Gong and Yuxia Liu

Application Research of Predictive Control in Heating System 343
Jinhong Wei, Heqiang Yuan, and Qianqian Wang

The Expert Controller for Dissolved Oxygen Control System in Wastewater Treatment 351
Jia Minzhi, Gao Binbin, and Jia Bo

Distributed Fuzzy Control System for Cotton Cloth Bleaching	357
<i>Xiaohong Guo and Congyan Yang</i>	
A Way to Diagnose the Rolling Bearing Fault Dealt with Wavelet-Packet and EMD	364
<i>Xiao-Feng Liu, Shu-Hua Wang, Yong-Wei Lv, and Xuan Lin</i>	
Exploring of Intelligent System for Oxygen Respirator Intelligent Training Check System Used in Colliery Salvation	370
<i>Mu-Qin Tian, Sen-Sheng Wei, Mu-Ling Tian, and Jun Guo</i>	
The Intelligent Calibrator of Mine Gas Sensor	377
<i>Jin-Yan Cao and Mu-Qin Tian</i>	
The Method of Power Saving and Study on Control System for Center Air-Condition Set	382
<i>Guojun Zhang, Yao Zhao, and Chaofeng Liu</i>	
Design of a Temperature and Humidity Monitoring/Sending System Based on SHT1 Temperature and Humidity Sensor	388
<i>Jun Guo, Mu-Qin Tian, and Mu-Ling Tian</i>	

Intelligent Image Processing

Research and Implementation for a Content-Based Video Playback System	392
<i>Zhang Lin, Duan Fu, and Li Gang</i>	
Research on Updating Land Use Database Dynamically by Remote Sensing Monitoring	399
<i>Xiuming Jia and Qiang Li</i>	
The Study of GPU-Based Parallel Hilbert Huang Transform	407
<i>Ningjun Ruan, Wen Zhang, ShengHui Yu, Kai Xie, and HuoQuan Yu</i>	
Designing and Recognizing Landmark for Positioning Underground Mine Vehicle	414
<i>Yu Meng, Hongmei Wu, Li Liu, and Wenhui Li</i>	
Researches on Combustion of Vacuum Switching Arc Based on Image Processing Method	420
<i>Huajun Dong, Yongquan Gan, Guiming Shi, and Jiyan Zou</i>	

Intelligent Scheduling

A Platform-Based Model for Automatic Planning and Corresponding Complexity Results	424
<i>Yunpeng Wu, Weiming Zhang, Zhong Liu, Jincai Huang, and Cheng Zhu</i>	

Research of Emergency Logistics Distribution Routing Optimization
Based on Improved Ant Colony Algorithm 430
Huijie Ding

Intelligent Signal Processing

A WSN-Based Pulse Wave Velocity Detection System For
Arteriosclerosis Monitoring 438
Deng Chen and Shao Chen

Condition Assessment Based on Gray Clustering with Cloud Whiten
Function for Power Transformer 446
Zheng Ruirui, Zhao Jiyin, Li Min, and Wu Baochun

Study on Intelligence Detection of Methane Gas Based on NDIR
Technology 454
Yong-zhi Liang, Xing-fu Rong, Bing Li, and Chun-tao Liu

Natural Language Processing

Research on Extracting Semantic Orientation of Chinese Text Based
on Multi-algorithm 461
Yanhui Zhu, Ping Wang, Zhihui Wu, and ZhiQiang Wen

Formal Semantics of Chinese Discourse Based on Compositional
Discourse Representation Theory 470
Qing-jiang Wang and Lin Zhang

Research of Chinese Word Sense Disambiguation Based on HowNet 477
Jingwen Zhan and Yanmin Chen

Nature Computation

Fast Quantum Algorithm of Solving the Protein Folding Problem in
the Two-Dimensional Hydrophobic–Hydrophilic Model on a Quantum
Computer 483
Weng-Long Chang

Parallelization of Wu’s Method with Multithreading 491
Hongbo Li and Suping Wu

The Influence of the Sectional Form of Labyrinth Emitter on the
Hydraulic Properties 499
Zhiqin Li and Lin Li

Neural Computation

- Neural Network and Support Vector Machines in Slime Flotation Soft
Sensor Modeling Simulation Research 506
Ranfeng Wang

Pattern Recognition

- The Research on Image Retrieval Based on Combined Multi-features
and Relevance Feedback 514
Shu-Juan Zhang
- Application of BP Neural Network in Drug Supervision Code Grade
Evaluation 521
Jing Zhao and Ye-li Li
- WordNet-Enhanced Dynamic Semantic Web Services Discovery 529
Li Chen, Zi-lin Song, Ying Zhang, and Zhuang Miao
- A New Scheduling Algorithm in Hadoop MapReduce 537
Zhiping Peng and Yanchun Ma
- The Kinetics Simulation of Robot Based on AutoCAD 544
Yu Ding
- An Improved Particle Swarm Optimization Algorithm 550
Dazhi Pan and Zhibin Liu
- An Improved Optimal Matching for Video Retrieval 557
Hu Shuangyan, Li Junshan, and Feng Fujun
- Improved Cue Fusion for Object Tracking Algorithm Based on Particle
Filter 564
Hui Li and Li Zhang
- A Study on Bus Routing Problem: An Ant Colony Optimization
Algorithm Approach 570
Min Huang
- The Research of ANN Forecasting Mode Based on Cloud Platforms 576
Jiang Xuesong, Wei Xiumei, Geng Yushui, and Wang Xingang

Rough Set Theory

- Trend Analysis of 2011 Chinas Car Sales Volume and Carbon Emissions
Based on Rough Set Theory 582
Li Yuansheng, Yang Yang, Fu Yanxiao, and Xu Xiangyang

The Application on Attribute Reduction by Using Bacterial Foraging Optimization and PSO Algorithm	590
<i>Wang Jianguo</i>	
A Novel Feature Selection Method for the Conditional Information Entropy Model	598
<i>Jing Ruan and Changsheng Zhang</i>	
Author Index	607

Optimization Model of Rotation Irrigation Channel Distribution with GA and FS

Weizeng Gao, Zhou Yu, and Guoyi Miao

Henan Institute of Science and Technology, HeNan Xinxiang 43000, China

Abstract. Optimal water distribution in irrigation district (ID) is an important method and countermeasure for reasonably restraining the water requirements, effectively increasing the water supply, and actively protecting the ecological environment etc. According to the genetic theory, it is essential to discarding the traditional binary system coding for solving water distribution. Considering GA sinking into the partial solution easily, a free search algorithm is introduced. It extend the search space for algorithms, enhancing the quality of gene coding, adjusting the discharge of canal , making every canal closed at the same time, decreasing the discarding water of canal, reducing the time of manual work. It has excellent theoretical and realistic effect on distributing water.

Keywords: Genetic algorithms, free search algorithms, the model of optimal distributing.

1 Introduction

Water distribution in Channel is one of the important contents of irrigation scheduling. Scientific and rational scheduling has greatly improving the utilization of irrigation water, reducing the waste water. The optimization of irrigation water can be divided into two models: one is aim at increasing benefits from irrigation water optimal distribution [1]; the second [2-4] is aim at minimizing water loss of irrigation channels. Influenced by cropping systems, rainfall, water irrigation, soil moisture, agricultural crop price; it is difficult to put use the first model in irrigation area for requiring many parameters that are confirmed hardly. In fact water distribution is confirmed by the smallest water losses, the shortest time in most current irrigation area according to actual situation. Being relatively simple, more convenient, few parameters required, the second model is more convenient for expand in irrigation area.

Lv hongxing and Song songbai have been researching it deeply, its coding according to binary system. It is inevitable to involve quantification error and consider balance the length of coding .Just as taking frequency sample from scattering continuous system would leading to lack fidelity, scattering continuous function would not to be true to the original, it could not reflect to the feature of its problem, because creating building blocks code is difficult. The longer of coding, the lesser of quantification. This would slow down the degree of convergence. So using natural coding, that is to say encoding the number of channel correspond to people thinking, Considering GA sinking into the partial solution easily , free search algorithms is introduced. It extend the search space for algorithms, enhancing the quality of gene

coding, adjusting the discharge of canal, making every canal close at the same time, decreasing the discarding water of canal, reducing the time of manual work.

2 Setting the Mathematical Model for Water Distribution

The model is aimed at solving the optimal schedule for water distribution according to the fixing quantity of flow, the changing times, and the different priority level. Supposing the quantity of flow of superior channel that has N subordinate channels is Q , the design flow and the actual flow in the subordinate channel is Q_{sj} , Q_j , ($j=1,2,\dots,N$), in view of demand for practical water distribution and decreasing the lose of water distribution in channel, variation of the actual flow would change 0.8~1.2 times to its design flow. T is rotation, t_j , t_{1j} , t_{2j} V_j respectively stand for the total time of water distribution, the beginning time, the ending time and demanding water yield in the j subordinate channel, $t_{2j} - t_{1j} = t_j$. The goal is to decreasing the water loss in superior and subordinate channel for water distribution.

Objective function

$$Z1 = \min(V_{su} + V_{sd}) \quad (1)$$

V_{su} , V_{sd} respectively stand for the total water loss in superior and subordinate channel.

According to document [5-6]

$$V_{su} = \frac{A_u L_u V_u Q_u^{-\mu}}{100} \quad (2)$$

A_u , μ respectively stand for coefficient of water permeability and index number of water permeability of the superior channel, L_u stands for the length of the superior channel, V_u stand for gross water requirements, Q_u stands for the actual flow in superior channel (m³/s).

$$V_{sd} = \frac{\sum_{j=1}^N f(A_j, L_j, V_j, Q_j^{-\text{mdj}})}{100} = \frac{\sum_{j=1}^N A_j L_j V_j Q_j^{-\text{mdj}}}{100} \quad (3)$$

N stands for the total number of channel, j stands for the ordinal number of the channel, A_j , mdj stand for coefficient of water permeability and index number of water permeability of the subordinate channel. L_j stands for the length of the subordinate channel, V_j stand for gross water requirements in subordinate channel, Q_j stands for the actual flow in subordinate channel (m³/s).

According to 1, 2,3, we have the formula as follow :

$$Z_1 = \min(V_{su} + V_{sd}) = \frac{\min(A_u L_u V_u Q_u^{-\mu} + \sum_{j=1}^N A_j L_j V_j Q_j^{-\text{mdj}})}{100} \quad (4)$$

A_u 、 μ_u 、 V_u 、 L_u 、 A_j 、 md_j 、 L_j 、 V_j is constant value, according to Z1, it is approximately considered that the more flow in channel, the less loss in gross water. if you want to decreasing the loss of gross water, you must be increasing the flow of water in channel.

The losing water in channel includes the transporting water loss and the abandoning water loss. So decision-maker expects the less time of water distributing, the good of results. The best result is all of channel will be closed at the same time. This will be able save manpower as well as decreasing water loss in channel. Because the parameter of channel is invariable, so the minimal time of water distributing is that:

$$Z_2 = \min \left(\max_{1 \leq j \leq N} t_{2j} - \min_{1 \leq j \leq N} t_{1j} \right) \quad (5)$$

$\max t_{2j}$ stands for the latest time of all channels, $\min t_{2j}$ stands for the first time of all channels. Every project for distributing water has its distributing time is that: $\max t_{2j} - \min t_{2j}$

Suppose at the time of day t , the latest channel begin to distributing water. There are $k-1$ channels still distributing water simultaneously. The latest finished time of $k-1$ channels is T_2 . So the goal of this function is to decreasing the time between these channels.

$$Z_3 = \min \left(\sum_{j=k}^N (T_2 - t_{2j}) \right) \quad (6)$$

Summing the different distributing time between all channels, our goal is to find the minimal time of all projects available.

N stands for the total number of subordinate channels, j stands for ordinal number of channels.

3 The Qualification of Model

(1) restriction on rotating time: suppose the longest allowed time for distributing water is T , the total time of all channels for distributing water in every project must less than T .

(2) restriction on the discharge of superior channels: the total discharge of subordinate channels must less than the discharge of superior channel at one time.

(3) restriction on the discharge of channels: the actual discharge of every channel must less than the designing discharge.

(4) restriction on the discharge: the time multiply the discharge equals to the total discharge of every channel, the extension of the discharge in every channel is $0.8 \sim 1.2$ times to its designing discharge.

4 Structure to Genetic Algorithm for Distributing Water

(1) confirming the variable and every restriction for design-making, the variable for design-making is Z_1 , Z_2 , Z_3 , restriction is mentioned above.

(2) set up the optimal model. According to objective, set up the function.

(3) set up chromosome code for every feasible solution, that is to describe the function with chromosome coding.

(4) set up the way to decode chromosome code, that is to translating the chromosome code to actual project for distributing water.

(5) set up the switch regulation from object function to individual fitness.

(6) set up the way for selective computation, crossover computation, mutational computation and so on.

(7) set up the initial condition: terminal evolving generation , initial population size, probability of crossover.

5 The Realization of Genetic Algorithms for Distributing Water

The theory of genetic algorithms is similar to different problem, so does the implement step. The difference is that men using different coding according to dissimilar problem. It is hard or easy to implementing genetic algorithms determined by the coding. The innovation in text is coding and mutation.

5.1 Coding

It is difficult to implement distributing water with binary system coding for having many distributing channels, extremely long code. In addition, with binary system coding to deal with problem is different to the way of men to solve problem. It will add difficulty to design program. So, the sequence of distributing channel will be taken to expressing the chromosome. Allele is n integer or n token.

Supposed the list files of type of the distributed channel is w_1 , assigning sequence number to every distributed channel, this list number is w_2 .

$$\begin{aligned} w_1 &= (v_1, v_2, v_3, \dots, v_n) \\ w_2 &= (1, 2, 3, \dots, n) \end{aligned} \tag{7}$$

V_1 - V_n corresponding to 1 – n one by one, using this coding:

T : 1 2 3 4 5 6n

The project will be expressed with T. Starting from V_1 , taking turn V_2 , V_3 , V_n , channel will be completely distributed.

5.2 Decoding

According to every project for distributing water, such as T: 1 2 3 4 5 6n, distributing water to k preceding channels. Suppose any one distributed channel is finished, checking in the discharge of $k+1$ channel, comparing the discharge of $k+1$ channel and the discharge in finished channel, if the former is larger, then distribute

water to k+1 channel, or waiting for another finished channel, add two finished channels discharge ,comparing again the discharge of k+1 channel and the sum discharge in finished channel, if the former is larger, then distribute water to k+1 channel, or waiting for another finished channel. According to the theory, channel will be distributed completely.

Without extra exterior information, the genetic algorithms will evaluate its chromosome with fitness function.

5.3 Fitness Function

There are three parameters to construct fitness function: Z_1 、 Z_2 、 Z_3 . Taking the max discharge will meet the satisfaction to Z_1 , Z_2 is selected as fitness function (the small, the good). Z_3 is reference fitness function. Z_3 will be used on condition that two project has the same value of Z_2 (the small, the good).

In distributing water, there are many restriction that will be met in coding process and decoding process expect rotation time and precedence level for distributing water. When decoding the chromosome, if one restriction could not be met, $f=Z_2 \times 2$, if two restriction could not be met, $f=Z_2 \times 4$ 。 Not only enlarging the difference between chromosome, The punishing computation will avail to convergence.

5.4 Design Gene Operator

Resembling to traditional gene algorithms, algorithms applied to this text is different in gene computation and evolutionary computation. Genome mutation applied to text is multiplicity, not only including mutation between genome that is exchange the position in every gene, and also mutation in genome that is change the discharge of distributed channel.

(1) design crossover operator The request for crossover operator is that: we will get two another practical project after crossing any pair distributing water project.

(2) design mutation operator A: Mutation in genome mutation has two functions in genetic algorithms: one is to provide variety for keep another operator work, the second is to act as search mutation. Mutation is always applying to single chromosome. There are many Mutation has many kinds such as inverse mutation, cross mutation, insert mutation. Inverse mutation is applied to the text. Inverse mutation is that selecting two genes in chromosome randomly and change its position one another. This will bring another new chromosome. For example, We will get T^x from T_x by changing the position between the fourth gene and the eighth gene in chromosome.

$$\begin{aligned} T_x &= (B C A | D E J H I | F G) \\ T^x &= (B C A | I H J E D | F G) \end{aligned}$$

B: mutation in genome Every channel changes its discharge within its allowed value. For any distributing water project, the key to problem-solving is to deal with disperse random value. Free Search(FS) algorithms are highly adaptable to the value.

Penev and Littlefair put forward Free Search(FS) algorithms recently[7]. In world, the animals such as horse or sheep has the innate qualities to looking for food. They cooperate each other and keep touch with each other within seeking process. The

animals has essential qualities : sense, mobile. Animal discern the road with sense in mobile. In seeking process, individual animal will change its searching way with its sense. It is similar to men to studying knowledge in world. In seeking process, individual change its searching way according to experience and knowledge. Individual could move to two directions: searching with little steps near to its position and searching with large step in global area. It is the innovation and advantage of Search algorithms. GA have the advantage of local searching, on the other hand , FS is in global searching. So applying GA to get relatively project, changing the discharge of channel with FS, we will get the best project for problem.

There are three ways for searching the discharge of channel: mutate at the same direction(increasing or decreasing at the same time), mutate at the different direction(the singular number increasing and the even number decreasing or the singular number decreasing and the even number increasing), mutate at random. Applying penalty function, searching form the designed discharge of channel, Searching steps is 0.0001 times of its designed discharge of channel.

5.5 Termination Conditions

The max number of generation of mutation will be set. The difference between the previous generation and the next generation in fitness value also will be set. The text takes the max number of generation is 2000.

Table 1. The rotational irrigation schedule of lateral canals in Fengjiashan ID (No.11 Branch Canal in North Main Channel)

Name of canal	area/hm ²	time/h	Name of canal	area/hm ²	time/h
1	82.93	138	13	53.33	75
2	33.93	47	14	213.33	281
3	102.33	132	15	26.13	36
4	112.53	98	16	183.3	267
5	34.13	48	17	35	47
6	96.67	161	18	23.67	33
7	46.07	65	19	171.2	264
New 7	74.27	102	20	201.33	333
8	62.73	98	New 20	36	49
9	28	40	21	109.33	171
10	60.53	89	22	33.8	51
11	52.6	94	23	7.93	11
12	103.33	155	24	47.87	76

Table 2. The results of distributing water

Name of canal	Timing begin (time/h)	Timing (time/h)	upLasting time (time/h)	Flow rate (m3/s)	Name of canal	timing begin (time/h)	timing (time/h)	upLasting time (time/h)	Flow rate (m3/s)
1	0	141	141	0.196	13	241.131	317.161	76.630	0.196
2	0	49.295	49.295	0.191	14	0	281.857	281.857	0.199
3	0	126.612	126.612	0.209	15	84.873	119.203	34.330	0.210
4	141	241.131	100.131	0.196	16	49.295	329	279.755	0.191
5	0	50.615	50.615	0.190	17	281.857	329	47.173	0.199
6	0	153.09	153.09	0.210	18	224.449	256.102	31.653	0.209
7	0	67.248	67.248	0.193	19	50.615	329	278.385	0.190
New 7	126.612	224.449	97.837	0.209	20	0	329	329	0.202
8	67.248	168.638	101.390	0.193	New 20	119.203	165.930	46.728	0.210
9	153.090	191.124	38.035	0.210	21	165.931	329	163.070	0.210
10	0	84.873	84.873	0.210	22	280.506	329	48.494	0.210
11	191.124	280.506	89.382	0.210	23	317.767	329	11.21	0.196
12	168.638	329	160.362	0.193	24	256.102	329	72.90	0.209

6 Experiments Results

Citing the data ShanXi province FengJiaShan irrigation area , It have a lot of branch channel in No.11 Branch Canal in North Main Channel. It is managed in two segment, one is called FaMen, the other is called ChengGuan. FaMen has branch channels:26, its designed discharge is 1.8m³/s. Discharge in each subordinate channel is 0.2m³/s. According to scheme for distributing water, rotation time is 14 d (336 h). Area and time for each channel is mentioned above.

7 Conclusion

(1) According as the genetic theory, it is essential to discarding the traditional binary system coding for solving water distribution. Being consistent with the way of thinking of the problem, decimal system is adopted. Using seriation of the canal for coding, it is enhancing the efficiency for coding and reducing the difficulty of programmer. It also make the result come close more to practice.

(2) Considering GA sinking into the partial solution easily , free search algorithms is introduced. It extend the search space for algorithms, enhancing the quality of gene coding, adjusting the discharge of canal , making every canal close at the same time, decreasing the discarding water of canal, reducing the time of manual work. It have excellent theoretical and realistic effect on distributing water .

References

1. Rao Vemuri, V., Cedeno, W.: New genetic algorithm for multi-objective optimization in water resource management. In: Proceedings of the IEEE Conference on Evolutionary Computation1, vol. 29, pp. 495–500. IEEE, Los Alamitos (1995)
2. Wang, M., Zheng, C.: Ground water management optimization using genetic algorithms and simulated annealing: Formulation and comparison. Journal of the American Water Resources Association 9, 519–530 (1998)
3. Morshed, J., Kaluarachchi, J.J.: Enhancements to genetic algorithm for optimal ground-water management. Journal of Hydrologic Engineering 51, 67–73 (2000)
4. Aral, M.M., Guan, J.: Optimal groundwater remediation differential genetic algorithm. In: International Conference on design using Computational Methods in Water Resources, CMWR, pp. 346–356 (1996)
5. Wong, H.S., Sun, N.-Z.: Optimization of conjunctive use of surface water and ground water with water quality constraints. In: Proceedings of the Annual Water Resources Planning and Management Conference, ASCE, pp. 408–413 (1997)
6. Ringler, C.: Optimal allocation and use of water resources in the Mekong River Basin: multi-country and intersectoral analyses. Development Economics and Policy Series, Frankfurt (2001)
7. Penev, K., Littlefair, G.: Free search-a comparative analysis. Information Sciences 122, 173–193 (2005)

Search Engine Based Maintainability Evaluation Measure of Web Site

Xiao-dong Wang and Jie Cui

The Telecommunication Engineering Institute,
Air Force Engineering University, East Fenghao road 1,
710077 Xi'an, China
hellowxd@mail.nwpu.edu.cn, watermelon_cj@163.com

Abstract. Evaluation measure of web site is an important tool for web mining and information release. Fatherly, maintainability is a notable merit among the evaluation. But, the measure which is able to stably attain standardized maintainability is absent in present. In fact, the merit is just implied in the search engines. When the spiders of search engine periodically visits web site to find new information and information update, both updating history and interval of web site page, which reflect the maintainability of web site, are recorded into search engine cache. By comparing latest web caches with current web pages, maintainability degree of web site can be calculated, and then maintainability of a web site is obtained too. The paper presents a search engine based maintainability evaluation measure of web site. Some experiments aim to test the measure illustrate the evaluation result and performance are credible and valuable.

Keywords: Web Site, Maintainability Evaluation, Search Engine.

1 Introduction

Web has spread to all over the world and becomes the most important media in human daily life. Every day, innumerable news, advertisement, video and other format information are continually released via web sites. If something wants to be known, the best method is to put it on a web site. But, there are hundreds of millions web sites on the web to be selected, how to select a suit web sites for information release become an issue.

In e-commerce domain, maintainability is an important merit for the selection, but it always is ignored or hardly attained. When a businessman selects web site to publish his advertisements, he usually prefers to one which is well maintained. But, how do they investigate the maintainability? Apparently, it is exhausted by repeatedly manual visitation. Especially, when number of candidates is very big, manual work becomes impossible to be carried out. Moreover, numerical maintainability evaluation is also useful reference for services promotion of web sites themselves. Hence, it is in agenda to design an automated maintainability evaluation measure.

In information retrieval domain, which is closely related to e-commerce, maintainability evaluation measure is also important. Ranking of search results, which consider

maintainability factor, will improve performance of search engines. By giving a search result rank according to information source quality, search engine will better satisfy user's long-term information requirement for similar topic.

Maintainability of web site may be directly reflected by its information update content volume and frequency. In other words, if web site's information (included in web pages) is updated sharply and frequently, it is regarded as being in good maintains status. Then, to detect information updating of a web site becomes an important precondition.

2 Design of the Evaluation Measure

In this paper, the evaluation of web site update mainly is implemented by a special detector, i.e. search engine. In order to find out useful information, Information Retrieval (IR) technology is developed by internet scientists in the last century. Today, many search engines based IR, such as altvista, overture, Google, Yahoo and so on, are available on the Web. Most of these search engines have smoothly visited web site on the web and recorded their variation for a long time. Thus, we may take use of search engines as detector to find out web site update.

There are two advantages to do so. Firstly, search engine has some effective mechanism, such as DNS cache, scalability, polite crawl, etc, to reduce negative affection. So detection will not disturb information services of web site. Secondly, because it need not design a special web site detector but using search engine, the cost of evaluation system is cut down obviously. Having detected update information, our maintainability evaluation measure calculates update content volume and interval respectively through comparison between web caches and current web pages. Finally, the two values are synthesized together into a normalization value.

We propose a novel measure to evaluate maintainability of web site by calculating its information update status, which is detected by search engine. The measure is called search engine based maintainability evaluation measure of web site. It contains three steps:

- Test if variance of web site occurs or not.
- Calculate content volume and interval of information update.
- Synthesize the two values together into a normalization value.

3 Implement of the Evaluation Measure

Update of a web site is measured on the basis of two factors, i.e. variance of web page content and interval update. For example the more content are replaced, the sharper update is considered, while the lesser interval of update, the frequenter update is.

Assume that evaluated web site is composed of text web pages. The whole process of search engine based maintainability evaluation measure of web site is illustrated in Fig. 1, it mainly includes following three steps.

3.1 Test the Variance of Web Site

Input home page URL of web site W waiting for evaluation into software system which is designed based our measure. In order to test whether the information of W is updated, two kinds of web page, i.e. current web page and latest web cache, are respectively collected from internet.

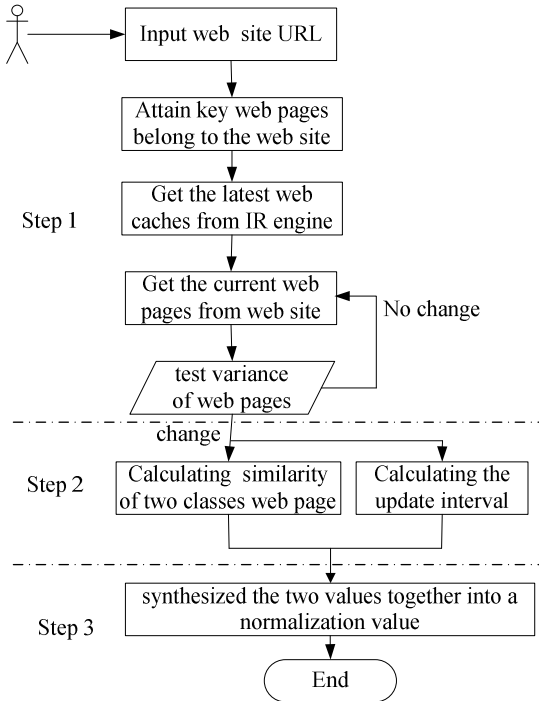


Fig. 1. The process of search engine based evaluation measure

1) Attain key web pages belong to web site. Start from home page of W , all of web pages under W will be traversed. In this course, each web page URL on the web directory sub tree is collected into a set, which is described as $\{u_1, \dots\}$. For the reason of simplicity, not all of web page but some key ones are considered, so a positive integer m is set as threshold. As long as the number of words on a page exceeds m , the page is looked as important as enough to be picked out. Finally, some web pages are selected, whose URLs compose a set U , $U = \{u_1, \dots, u_i, \dots, u_n\}$, where n is total of pages selected, $1 \leq i \leq n$. The bigger n is, the more precise following calculation result is. On the other hand, a big n means long calculation time.

2) Get the latest web caches from search engine. As following mentioned (Related Background), web caches of search engine record history information of web pages. So a web caches set WC , where $WC = \{WC_1, \dots, WC_i, \dots, WC_n\}$ and WC_i corresponding

to u_i in U , are downloaded. Treat the words in WC_i as a sequence and use MD5 algorithm to calculate the signature of every web caches in WC . The signatures are put into a set WC^{MD5} in turn, here $WC^{MD5} = \{WC_1^{MD5}, \dots, WC_i^{MD5}, \dots, WC_n^{MD5}\}$.

Function of query on cache of search engines is used to get cache pages. Most of search engines provide such function. For example, Google has “cache” command. If user import “cache:” plus a URL of a web cache into browser address bar, the page cache will be presented. Please note that a WC_i always can not be smoothly downloaded by single time. There are three possible reasons. The first one is that the web page corresponding to u_i has been removed. The second one is that the web page corresponding to u_i is so new that search engine spiders have not recorded. The third one is communication fault. So several times attempt should be done.

When WC is downloaded, some text pretreatment will be done. Stop words are eliminated, following with word stemming. In succession, page WC_i is represented into vector of TF/IDF form, i.e. $\overline{WC}_i = \{(t_1, w_{c_{1,i}}), \dots, (t_j, w_{c_{j,i}}), \dots, (t_N, w_{c_{N,i}})\}$, where t_j is a key term which come from training corpus, N is dimension of vector, $1 \leq j \leq N$, $w_{j,i}$ is the weight of t_j in WC_i .

Extract the create time of WC_i , namely t^{WC_i} , from web caches in turn, then form a set $\{t^{WC_1}, \dots, t^{WC_i}, \dots, t^{WC_n}\}$. If WC_i download fail, then let $\overline{WC}_i = \{(t_1, 0), \dots, (t_j, 0), \dots, (t_N, 0)\}$, $WC_i^{MD5} = -1$, $t^{WC_i} = t_{cycle} \cdot t_{cycle}$ is the cycle time of search engine spider. Usually t_{cycle} is a constant and set by web site administrator.

3) Get current web pages from web site. Use U to download current web pages of W . These pages compose a set WP , $WP = \{WP_1, \dots, WP_i, \dots, WP_n\}$. Like WC , the signature set WP^{MD5} of WP , $WP^{MD5} = \{WP_1^{MD5}, \dots, WP_n^{MD5}\}$ and tf/idf vector $\overline{WP}_i = \{(t_1, w_{p_{1,i}}), \dots, (t_2, w_{p_{j,i}}), \dots, (t_N, w_{p_{N,i}})\}$ are gotten. If WP_i is deleted, $w_{p_{j,i}} \equiv 0$.

Get current local system time t_{cur} .

4) Test the variance information of web site. If all of $WC_i^{MD5} = WP_i^{MD5}$,

Then it is regarded that there is no variance information, and system goes to 3) after a certain interval. Otherwise, information updating of U is regarded to already occur.

3.2 Calculate the Content Volume and Interval of Information Update

In this step, the maintainability of U is evaluated form two aspect, namely volume and interval of information update. The former means how much information is updated, and the later how long time is.

1) Calculating volume of information update. Use formula (1) to calculate d_U .

$$d_U = \sum_{i=1, WC_i^{MD5} \neq WP_i^{MD5}}^n \text{sim}(\overline{WP}_i, \overline{WC}_i) \quad (1)$$

Where $\text{sim}(\overline{WP}_i, \overline{WC}_i)$ is the similarity of \overline{WP}_i and \overline{WC}_i , calculated by formula (2):

$$\overrightarrow{sim}(WP_i, WC_i) = \frac{\sum_{j=1}^N wp_{i,j} \times wc_{i,j}}{\sqrt{\sum_{j=1}^N wp_{i,j}^2} \times \sqrt{\sum_{j=1}^N wc_{i,j}^2}} \quad (2)$$

2) Calculating interval of information update. Calculation of the average interval \bar{T} is as following formula:

$$\bar{T} = \sum_{i=1, WC_i^{MD5} \neq WP_i^{MD5}}^n (t_{cur} - t^{WC_i}) / k \quad (3)$$

k is the total of u_i in U , which $WC_i^{MD5} \neq WP_i^{MD5}$, $1 \leq i \leq n$.

3.3 Synthesized the Two Values Together into a Normalization Value

In this step, volume and interval of information update will be synthesized to value M , $0 < M < 1$, as follows:

$$M = \beta \cdot (1 - \frac{d_u}{n}) + \gamma \cdot (1 - \frac{t_{cycle}}{\bar{T}}) \quad (4)$$

Where β, λ are respectively the weigh of d_u and \bar{T} , $\beta > 0, \gamma > 0$, $\beta + \gamma = 1$.

In order to improve the calculating precision, the whole process should be repeated several times, and an average Normalization value corresponding to maintainability of U is finally given to user.

Although the measure mainly process text web pages now, it can be easily extended to multimedia web pages by using voice recognition techniques, image/ video processing techniques. To extend it to multimedia web page, only two jobs should be done. First, in step 1(mention in A), some suitable expresses for image or video will take the place of TF/IDF vector. Another is to replace formula (2) by image or video similarity calculation. Therefore, the measure will be still feasible.

4 Evaluation Experiments and Results

In this section, we build up software based on the measure to implement some experiment missions. Two famous businesses web site and two university ones are selected to the test. For the reason business benefit, the names of business portal are hidden. The software is programmed with IDE VC++6.0, WinInet API, and Google cache pages. The measure is tested from two aspects, i.e. maintainability and performance.

4.1 Results for Goal 1: Maintainability

In this experiment, we deeply test the measure from two hands with respect to the maintainability. On the one hand, we will compare maintainability degree of websites to find out if they agree with humanity feeling. On the other hand, a specific web site is picked out to analyses its maintainability changing tendency in a period time.

1) Maintainability degree comparing between websites

At every test point, we implement our software at the last day of three month in 2011, and attained the result as table 1.

Table 1. Maintainability Degree Comparing.

<i>website</i>	<i>Jan. 31</i>	<i>Feb. 28</i>	<i>Mar. 31</i>
bussiness website1	0.33	0.41	0.36
bussiness website2	0.42	0.37	0.39
www.nwpu.edn.cn	0.23	0.04	0.27
www.afeu.cn	0.07	0.01	0.03

From the table we know that the maintainability of business websites is obviously higher than education ones' a lot. The result is consistent with fact.

2) Maintainability Changing Tendency Experiment

www.nwpu.edu.cn is selected to the tendency analysis from 2010 Oct. to 2011 Mar. The test points also are the last day of every month. Test result is illustrated in Fig. 2.

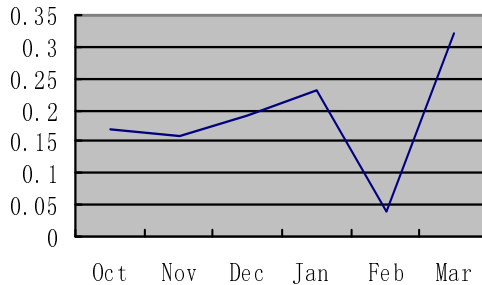


Fig. 2. The maintainability changing tendency of www.nwpu.edu

From figure 2, we know that the maintainability form Oct to Jan is average, Feb is the minimal and Mar is a value peak, the analysis about the tendency as following. We know the time form Oct to Jan is normal time of the university, the information updating is gently. Feb is Chinese traditional festival “spring festival” in 2011, all of staff and students are in vacation, so that little information is updated. That means the maintainability must be the lowest ebb. Since Mar is the first month of a new term, much information is released in it.

In conclusion, maintainability changing tendency, which is attained with our measure, is also rational.

4.2 Results for Goal 2: Performance

The calculating time is main performance merit of our measure. In this experiment, we recorded the time exhaust in the evaluation, and the result is satisfied basically. The average time is 22.6 min for a website including about 20 thousand pages, that is benefited a lot form the using of multi-thread technique in the programming.

According to farther exhaust analysis, the main time is spent on the TF/IDF vector calculation. It is certainty that some optimized algorithm ought to be adopted in the future. In addition, network traffic also affects performance directly. Another interesting thing is that toward different website the time exhaust is very different even if the occasion, network traffic, workload are coincide. We think it must be led by the search policy of the engine.

5 Conclusion and Future Work

Form the evaluation result and performance attained in the experiment, we can draw a conclusion the measure is credible and valuable. Our measure can improve performance of search engines. It provides a new parameter, about information source quality, to search engine result rank, which will better satisfy user's long-term information requirement for similar topic. Software based our measure may be seamlessly embedded in search engine. But a flaw also ought to not be neglect, namely its works depend on the performance of search engines. In the future day, following works should be done.

Results of evaluation can directly supply references to advertisers before they contract with a web site. Web site also can use this tool to evaluation its information update rate in order to promote their services. So some commercial software based on this measure will be gotten onto market. Now, our measure is mainly used to evaluation text web page. But, today most of pages include various elements. So the more type of web elements ought to be supported. This job is feasible and ought to be done as soon as possible is next step work. The measure is also useful to other web applications today and future day. To some degree, maintainability means novelty. We may use similar measure to calculate how much the service shift which a web services provide in a period of time. From the result, the novelty of services is very clear. So using the thinking of this measure into Web service evaluation is a challenge job to be implemented. Search engine and web site are the most foundational services on the web, so our measure can be integrated into many applications.

6 Background and Related Work

In this section, we describe some knowledge about search engine technique, and explain why it can get information of a web site and record web pages variance. Some related work of web site evaluation will be introduced.

Each web page has many hyperlinks. Those hyperlinks link to other ones, which also has hyperlinks. All hyperlinks compose a network which connects almost every web page on the web. Spider system of search engine goes along the web to download web pages into its database step by step. While user input some terms into search box of search engine, web pages which include terms in database will be returned back. Because web pages are not invariable, the database must be update in time, namely to download modified web pages again. When web page database of search engine is updated mainly bases two factors. One is that some new information is added and is enough to cause search update. The other is that update period of spider system is

reached. The update period can set by either web site or search engine. Usually, overdue web pages are not immediately deleted and will be preserved for a time. These web pages are called web cache, which save history of web pages variance.

There already exist many methods to evaluate web site. But most of them prefer to permanence of web site self, such as speed of data transmission, security, navigation efficiency, etc [1-2], and maintainability of web site often is ignored. After all, essential function of web site is to release information. If information is updated inopportunately, the web site is considered to unhealthy or under careless maintenance. In a word, information update directly reflects maintainability of web site. Based on the principle, we design a method to evaluate a web site.

As mentioned in the Part 1, web sites play an important role in e-commerce and IR, so how to evaluate a web site is important. Many jobs have been done. But Most of evaluation measures focus on performance or visitor action of web site, such as usability, flux, click rate, navigation efficiency, security, data transmission speed, and so on [3-5].

Differing from others, this measure evaluates maintainability of web sites by calculating its information update status. Implementation mainly takes use of function that searching engine periodically collects web pages on the web, so detector of system need not to be redesigned and the cost is cut down notably.

Moreover, some favorable characters of search engine are inhered, for instance high-performance, scalable, distributed, DNS cache, and so on. Although our method mainly process text web pages now, it can be easily extended to multimedia web pages by using voice recognition techniques, image/video processing techniques.

Acknowledgment. This work is part of The Research of Semantic Information Resource Detecting Technique Based on Ontology. The research is supported by the province natural science foundation of Shaan'Xi under grant NO.2010JM8004 and 2010 doctor research initial funding of AFEU.

References

1. Barnes, S., Vidgen, R.: WebQual: an Exploration of Web Site Quality. In: Proceedings of the Eighth European Conference on Information Systems, Vienna (2000)
2. Loiacono, E., Watson, R., Goodhue, D.: WebQual(tm) : A Web Site Quality Instrument. In: American Marketing Association : Winter Marketing Educators Conference, Austin, Texas (2002)
3. Dragulanescu, N.G.: Website Quality Evaluations: Criteria and Tools. Information & Library Review (2002)
4. Bevan, N.: Usability Issues in Web Site Design [EB/OL] (2007), http://www.lboro.ac.uk/research/Husat/inuself3_web_paper.html
5. Weilin, Y., Xiangmin, Z.: To Study on the Quality Evaluation Method Based on Users Satisfying at Website, Value Engineering (2008)

Research on Modeling Plant Growth in Greenhouse Based on Virtual Plant Technology

Tang Wei-dong¹, Li Jin-zhong¹, and Hu Xue-hua²

¹ School of Electronics and Information Engineering, Jinggangshan University,
Ji'an, Jiangxi Province, China

² School of Life Science, Jinggangshan University,
Ji'an, Jiangxi Province, China
twd_1974@126.com

Abstract. Modeling plant growth is important to further study on the plant development process, especially for digital agriculture field. Virtual plant model can help to observe the plant growth process and discover some recondite rules, which could be constructed based on large numbers of experimental data during the plant development. The growth model can be obtained after dealing with experimental data, while the changes including topological structure and organ morphology during the plant growth could be described using virtual technology. Then the virtual plant development model can be expressed with the information fusion and reconstruction method. The experimental results showed that the presented model was feasible and effective on simulating plant growth on computer, which also greatly demonstrated its valuable evidences in predicting the plant growth rules under different environmental conditions.

Keywords: Virtual plant technology, Development model, Plant growth.

1 Introduction

After rapid development of computational information technology and its applications in different areas such as digital agriculture, using the traditional or existing plant growth models could hardly keep up with the modern agricultural development, while the interested areas related to modern agriculture such as virtual plant technology has been caught up great attraction by some native and foreign researchers. Owing to the important position in the virtual agriculture and modern agricultural technology, virtual plant technology and its applications in modern agriculture have been greatly popularized. Researches in countries such as France and Netherlands constructed virtual plant models of maize and cotton and so on, using L system or fractal method or AMAP way according to plant development and experienced data, which could get hold of the plant growth and development to some extent[1-4]. Due to studying on this field a little later for Chinese researchers, who mostly concentrated on construction methods and visualization of virtual plant of the main crop such as soybean and rice plant and so on[5-8]. Though methods about how to model virtual plant had been proposed by some researchers, such as L-system and its extension theories, structural and functional model, and so on, by which it used to simulate

morphological changes of plant, but the development process of plant can not be really known because of simplifying the modeling so as to usually ignore the effect on plant made by outer environment. Then how to construct the virtual plant development model fitting to the growth law of botany is to be the interested problem within this field. In addition, researches on virtual plant model are generally focused on farm crop, while the crops in greenhouse have not been showed great interest by researchers, especially it is hardly paid attention to how to construct the virtual plant model in greenhouse based on the plant development law. Thus, studying on the main crop cucumber in greenhouse as a case, this paper introduces such method as using virtual plant technology to obtain the virtual cucumber model based on the existing growth models, which will offer decision-making and theoretic guidance in modern agricultural production and management.

2 Materials and Experiments

Generally, the experimental object is the Netherlands cucumber randomly selected from modern greenhouse in Jiangsu Ruijing Agricultural Technology Garden. The experiments operated in spring and autumn. During the period of cucumber began to burgeon to mature randomly selected plants with similar growth to do the destructive measure per week. Marking the data from spring and autumn as S_1 and S_2 . The S_1 used to model construction and parameters estimation while S_2 used to model test. Separated various organs in order to observe and measure. The parameters including stem length, the number of internode of plants leaf area index, the fresh weight and dry weight of organs. Then taking the average corresponding data as the observational data. Using the electronic balance with the accuracy of 0.001g to measure the fresh weight. Then they will be placed the oven in 120 °C for 30 minutes, then cool down to 80 °C bake for 48 hours. After drying to constant weight measure the dry weight of organs separately. Meteorological data collected from the local observatory.

3 Virtual Plant Technology

3.1 Growth Model Construction

Usually, it could describe the plant growth using the stem height index, while the stem growth of plant is related to the change of inter-node, for the inter-node numbers and their growth rate change with their growth age. Owing to the cucumber plant development affected by temperature, it could use GDD (Accumulative Growing Degree Day) as parameter to construct the dynamical simulation models of stem height and growth age as well as leaf area index after the experimental observation during the period of spring.

Then it could construct the model of stem height changing with the GDD with SPSS (Statistical Program for Social Science) packet on computer. According to the results from experimental observation, to get the simulation model, it would select such curves as Growth curve, Exponential curve, Logistic curve and S curve as well as other similar simulation modeling.

3.2 Description Model of Plant Growth

Due to the morphologic and structural changes of cucumber since its burgeon time, which including topological and morphological change of plant or organs. It could convert the plant growth information to regulative visual graphic one in order to reshape the plant development process on computer.

It could construct the dynamic model of organ growth changing with the GDD with SPSS(Statistical Program for Social Science) packet on computer. According to the experimental results, to get the simulation model, it would choose such curves as Growth curve, Exponential curve, Logistic curve and S curve as well as other similar simulation modeling. For example, if marking the morphology of inter-node organ as y , the value of GDD is x still, f is the relationship between them, then it can conclude the following function through experiments.

$$y = f(x) \quad (1)$$

3.3 Data Processing

The data processing in relation to plant growth generally includes information fusion and reconstruction. For the systematic change process during plant development, to realize the visualization of plant growth on computer, it has to effectively complete the information fusion between growth unit and organs class according to the properties of plant growth in different stages. Then with the method or technology of D-S theory and fuzzy one, it could combine and correlate the information of topologic structure and organ morphology according to the description models of plant development in order to get the growth information[6]. Furthermore, to simulate the process of plant growth on computer, it has to reconstruct the information related to plant growth with the method of virtual plant technology based on information fusion, and it is important to express the plant development process in numerical information so that it can transfer the abstract information of plant growth to concrete digital one. While reappearing the plant growth course in reality, it has to combine the information embodying in plant growth mechanism with that of topological structure and organ morphology of plant when making morphological structure of plant based on the theory of modeling virtual plant. Thus, it has to reconstruct correlated growth information according to topological change rule of plant and morphological change law of organs when simulating the change process of morphological structure of plant on computer. Fig. 1 indicates the process of various kinds of information fusion and reconstruction. On making the reconstructing information implying the law of plant development and reappear on computer, the information reconstructing work mainly includes both information mapping and information normalization process[8]. When doing information mapping process, using morphological information including in the parameter database of plant growth abstracted from experimental results based on the existing morphological model of plant, the topological structure information of plant and morphological information of organs should be mapped to the corresponding geometric graphic information such as points, lines, surfaces, etc. based on graphic technology, then the information could be organized and saved in some data structures for later use. So far as information normalization process is concerned, it generally deals with the geometric graphic information at various development stages

of organs, and then the normalizing graphic database of organs can be constructed based on such methods as Nurbs technology and graphic function models of organs. In additional, not only the interaction between plant and outer environment, but also the changes of topological structure, organ morphology, physiology as well as ecology during the plant development are always correlated and reacted with one another. To obtain the systematic information in relation to plant growth, it generally has to connect plant growth information with environmental information.

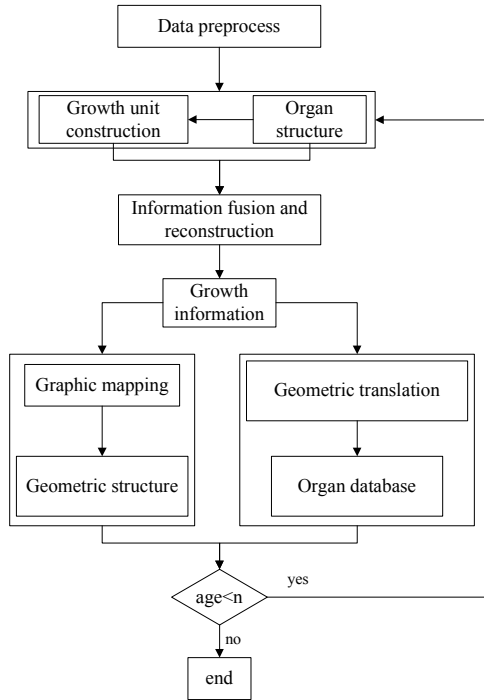


Fig. 1. Diagram of data processing

3.4 Modeling Virtual Plant

Usually, topological structure of plant changes with plant sorts and its development stages. In order to get the changing rule of the topological structure of plant, it has to obtain the changing rule of the topological structure of plant that could indicate the properties of different growth stages. For example, Fig. 2 shows the changing process of the topological structure of plant. The topological structure of plant could be illustrated in such terms as L-system, especially, the Open-L system method used to modeling the plant growth in order to embody the correlation between plant growth and outer environment[6]. Moreover, the morphological change of plant organ is related with plant and organ sorts, and with organ development phases. The change of organ morphology could be described by some geometric parameters of plant organ. In view of similarities of the same kind of plant organ, the organ morphology could also be described using the organ variables, thus, the change of every organ morphology

could also be described by the corresponding organ variables. To get the topological change rule of plant, according to the characteristics of plant growth at every development stages, it should firstly obtain the correlation between morphology and physiology through periodical experiment. Using growth parameters data-base representing the changes of plant topologic structure and organ morphology during the plant development, it could make some quantitative analysis of morphological change of organs during their development, and then the morphological changing model is concluded by the topological change rule of plant.

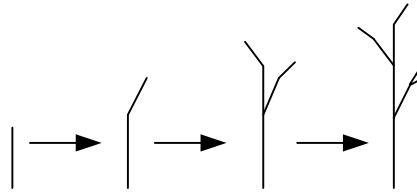


Fig. 2. Diagram of processing topological change of plant

To continuously simulate the plant growth process dynamically, it should construct the digital models of virtual plant, which including both geometric model and displaying one related to morphology of plant. On constructing the geometric model, it has to use the database such as plant growth parameters and organ graphics based on information reconstruction and rule of the topological change of plant. Furthermore, the displaying model can be constructed by using the continuity of plant growth process and combining models in relation to texture and shadow as well as light of organs. Fig. 3 demonstrates the constructing diagram of digital of virtual plant.

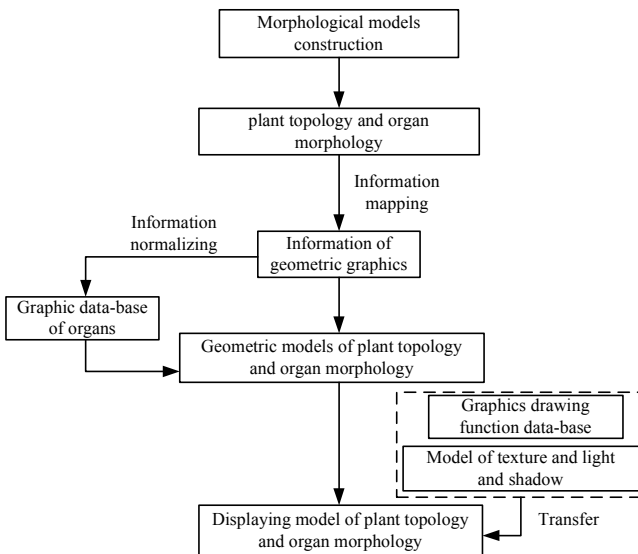


Fig. 3. Diagram of modeling virtual plant process

4 Simulation and Analysis

With Visual c++ as a developing tool, combining with the open graphic library which provides some important database of main functions and utilization ones, this paper developed the system on virtual plant growth. Taking both characteristics of plant growth and influences of outer environment into account, the system could comprise the dialog box of basic attributes configuration, which aims to modify such attributes of organs as kinds of leaf and angle between leaf and stem, and so on, in the meantime the outer environment variables could be got through loading files containing meteoric data. After modifying some attributes and conditions of outer environment, it would get such simulation results as in Fig. 4, which showed the normal development of plant under suitable conditions. In terms of the plant development, the morphological structure of plant, such as stem and leaf, always changes conforming to some laws with the effects of outer environment including temperature and fertilization and others suitable for the plant development.



Fig. 4. Simulation results of virtual plant development

5 Conclusions

Using the conventional growth models and modern computer and dealing tools, this paper introduces the virtual plant modeling method, and applied in cucumber in greenhouse as a case. Analyzing the change rule of plant topology and organ structure after processing experimental results and information fusion, it presented the morphological model of plant, then introduced the method of modeling virtual plant based on information fusion for simulating the plant growth process conforming to the law of botany. The description models about plant topology and organ morphology could be expressed based virtual plant technology, then, the virtual models based on information fusion and reconstruction could be constructed. From the simulation results, it showed that the models of plant growth not only represent the botany law to some extent, but it will propound important basis for further studying the effect of various environmental factors on plant growth. Additionally, accounting for the complexity of plant sorts and their morphological structure, the correlation between plant and outer environment could be obtained depending on additional large experiments, such as the law of change of geometric structure and physiologic characteristics under various

environmental factors. Therefore, it has to consider lots of factors when applying the method to specific plant, which could be the difficult problem in modeling virtual plant, and would have to take much more time to settle down after continuous experimental observation and accumulation of work.

Acknowledgments. We thank the support of this work by Foundation for Youths of Jiangxi Provincial Department of Education of P.R. China (item number: GJJ09591) and the support of Doctoral Foundation by the Administration of Education of P.R. China (item number: 20060299003). The project sponsored by the scientific research foundation for the doctoral scholars of Jinggangshan University.

References

1. Jallas, E., Sequeira, R., Martin, P., et al.: Mechanistic Virtual Modeling: Coupling a Plant Simulation Model with a Three-dimensional Plant Architecture Component. *Environmental Modeling and Assessment* 14, 29–45 (2009)
2. Koes, R.: Evolution and development of virtual inflorescences. *Trends in Plant Science* 13, 1–3 (2008)
3. Prusinkiewicz, P.: Art and science for life: designing and growing virtual plants with L-systems. *Acta Horticulturae* 630, 15–28 (2004)
4. de Moraes Frasson, R.P., Krajewski, W.F.: Three-dimensional digital model of a maize plant. *Agricultural and Forest Meteorology* 150, 478–488 (2010)
5. Weidong, T., Pingping, L., Zhangping, L.: Technology of constructing virtual plant system including model coupling. *Transactions of the Chinese Society for Agricultural Machinery* 39, 94–98 (2008)
6. Weidong, T.: Study on virtual plant technology based on plant growth model——dominant plant-reed in the beigu-mountain wetland as a case. Jiangsu University, Zhenjiang (2007)
7. He, H.-j., Yang, H.-y., Tang, J.-j., et al.: Design and Realization of Virtual Growth Visualization System for Rice Plant. *Journal of System Simulation* 21, 4393–4396 (2009)
8. Tang, W., Liu, C., Li, P., et al.: Virtual plant model based on open-L system and recursive expression. *Transactions of the Chinese Society for Agricultural Machinery* 40, 167–170 (2009)

Credit Assessment with Random Forests

Lei Shi, Yi Liu, and Xinming Ma

College of Information and Management Science, HeNan Agricultural University,
Zhengzhou 450002, China

sleicn@126.com, 13592621857@139.com, xinmingma@126.com

Abstract. Because the credit industry has a lot of bad debt problems, credit assessment has become a very important topic in financial institutions. Recent studies have shown that many algorithms in the fields of machine learning and artificial intelligence are competitive to statistical methods for credit assessment. Random forests, one of the most popular ensemble learning techniques, is introduced to the credit assessment problem in this paper. An experimental evaluation of different methods is carried out on the public dataset. The experimental results indicate that the random forests method improves the performance obviously.

Keywords: Random forests, Ensemble learning, Credit assessment.

1 Introduction

With the growth in electronic commerce, the amount of trade conducted electronically has grown extraordinarily and the credit industry has experienced two decades of rapid growth with obvious increases in auto-financing and credit card debt. Recently, everyone can be easy to apply for the credit cards in the fields of electronic commerce. However, many financial institutions suffered heavy loss from a steady increase of customers' defaults on loans [1]. Thus, credit assessment has become an increasingly important area for financial institutions.

Recently, many credit assessment methods have been widely studied in the areas of statistics, machine learning and artificial intelligence. The advantages of credit assessment include reducing the cost of credit analysis, enabling faster credit decisions, predicting the credit card risk and freezing the credit usage. Numerous statistical techniques have been used for improving the performance of credit assessment. The linear discriminant analysis and the logistic regression are the two most popular parametric statistical methods for credit assessment. In [2], linear discriminant analysis is introduced as one of the first parametric statistical methods for credit assessment. In [3], the logistic regression approach is used for credit assessment to overcome the deficiencies of linear discriminant analysis in the application of credit assessment. Because parametric statistical methods are of low performance in the credit assessment, many algorithms in the fields of machine learning and artificial intelligence are proposed for efficient credit card risk prediction, such as Support Vector Machines (SVM) [4, 5], genetic programming

[6, 7], artificial neural networks [8], k-nearest neighbor classifier [9], decision tree [10, 11], etc.

Ensemble learning is learning algorithm that trains a set of component classifiers and then combines their predictions to classify new patterns [12]. Ensemble learning is one of the four current directions in machine learning and has attracted more and more attention in the machine learning and data mining communities. Random forests developed by Leo Breiman and Adele Cutler is an ensemble classifier that consists of many decision trees and outputs the class that is the mode of the class's output by individual trees [13, 14].

With the growth of the credit industry and the large loan portfolios under management today, credit industry is actively developing more accurate credit assessment methods. In this paper, the random forests, one of the most popular ensemble learning techniques, is proposed to predict credit card risk. The experiments are conducted on the public dataset. The experimental results indicate that the random forests technique improves the performance obviously.

The remainder of the paper is organized as follows. In section 2 we will introduce the random forests technique in detail for the sake of further discussion. In section 3 we will report the experimental evaluations. The conclusions are given in section 4.

2 Random Forests

Random forests is an ensemble of classification or regression trees, induced from bootstrap samples of the training data, using random feature selection in the tree induction process [13]. Fig. 1 presents a general architecture of random forests, where n is the number of trees in random forests and y_1, y_i, y_n , and y are class labels [15].

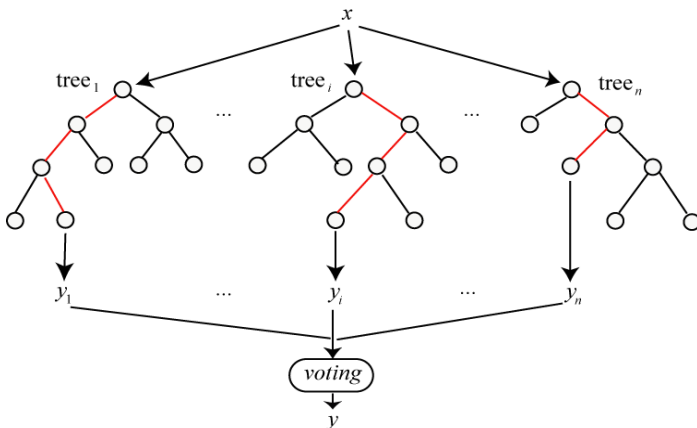


Fig. 1. A schematic view of random forests

The strategy of random forests is to select randomly subsets to grow trees, each tree being grown on a bootstrap sample of the training set. For example, bootstrap sample S_i contains m randomly drawn patterns with replacement from the original

training sample and then a decision tree t is built on the bootstrap sample S_t . Then, the random forests combines a collection of T classification or regression trees. For each pattern, each individual tree votes for one class and the forests predicts the class that has most votes, i.e. the pattern is classified into the class having the most votes over all T trees in the forest. Random forests generally exhibits a substantial performance improvement over the single tree classifier such as C4.5. It yields generalization error rate that compares favorably to Adaboost, yet is more robust to noise [15, 16]. Compared with other ensemble classifiers, random forests has the advantages as follows: the variance reduction achieved through averaging over learners and randomized stages decreasing correlation between distinctive learners in the ensemble.

3 Experimental Evaluation

3.1 Experimental Datasets

In this section, we present an experimental evaluation of the random forests technique for the credit card risk prediction. We have employed a dataset from the UCI dataset repository, i.e., the German credit dataset [17]. In the German credit dataset, the number of patterns is 1000, the number of attributes to describe a German is 20 and the number of classes is 2.

3.2 Performance Measures

To analyze the performance of classification, we adopt the Accuracy. As shown in Table 1, four cases are considered as the result of classifier to the pattern [18].

Table 1. Cases of the classification for one class

Class C		Result of classifier	
		Belong	Not belong
Real classification	Belong	TP	FN
	Not belong	FP	TN

Using these quantities, the performance of the classification can be evaluated in terms of Accuracy.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

3.3 Results and Discussion

To evaluate the performance of the random forests technique, C4.5 and decisionstump are used as benchmarks for comparison in the experiments. Performance is evaluated by 10-fold cross validation. We split each dataset into ten parts. Then we use nine parts for training and the remaining tenth for test. We conduct the training-test procedure ten times and use the average of the ten performances as final result.

In the first experiment, we compare the random forests method with other popular algorithms. The number of trees in random forests is set as 10.

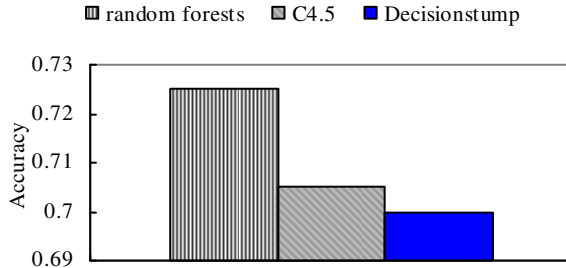


Fig. 2. Comparison of the Accuracy of prediction on German credit dataset

Figure 2 shows the prediction results of various techniques in terms of Accuracy value on German credit dataset. The prediction Accuracy of random forests is 72.5%, which is approximately 2.5% higher than that of decisionstump, 2.0% higher than that of C4.5.

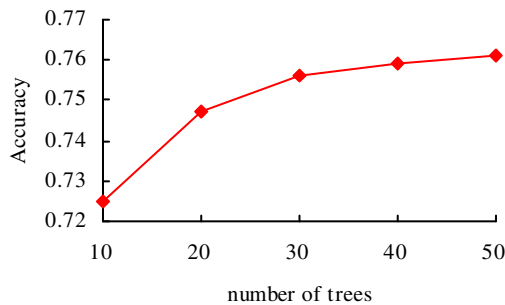


Fig. 3. Accuracy of the random forest on the German credit dataset

In the second experiment we investigate the dependence of the classification accuracy with the number of trees in random forests. We adjust the number of trees in random forests from 10 to 50 with step 10 respectively to create a wide range of settings and analyze the sensitiveness of the performance of random forests to number of trees. Figures 3 shows how the number of trees in the random forests affects the performance of the prediction on the German credit dataset.

4 Conclusion

Credit industry has been rapidly developed over past few years. Because the credit industry has lots of bad debt problems, the credit assessment has become a very important topic. This paper has applied the random forests approach to credit assessment and reported a comparison study with other popular methods. The experiment results indicate that the random forests method is an effective classification tool for credit assessment.

References

1. HSN Consultants Inc.: The Nilson report, Oxnard, California (2007)
2. Reichert, A.K., Cho, C.C., Wagner, G.M.: An examination of the conceptual issues involved in developing credit-scoring models. *Journal of Business and Economic Statistics* 1(2), 101–114 (1983)
3. Henley, W.E.: Statistical aspects of credit scoring. Dissertation. Springer, Milton Keynes (1995)
4. Huang, C.L., Chen, M.C., Wang, C.J.: Credit scoring with a data mining approach based on support vector machines. *Expert Systems with Applications* 33(4), 847–856 (2007)
5. Ligang, Z., Kin, K.L., Jerome, Y.: Credit Scoring Models with AUC Maximization Based on Weighted SVM. *International Journal of Information Technology and Decision Making* 8(4), 677–696 (2009)
6. Abdou, H.A.: Genetic programming for credit scoring: The case of Egyptian public sector banks. *Expert Systems With Applications* 36(9), 11402–11417 (2009)
7. Ong, C.S., Huang, J.J., Tzeng, G.H.: Building credit scoring systems using genetic programming. *Expert Systems with Applications* 29, 41–47 (2005)
8. West, D.: Neural network credit scoring models. *Computers and Operational Research* 27, 1131–1152 (2000)
9. Henley, W.E., Hand, D.J.: A k-nearest neighbor classifier for assessing consumer risk. *Statistica* 44(1), 77–95 (1996)
10. Davis, R.H., Edelman, D.B., Gammerman, A.J.: Machine learning algorithms for credit-card applications. *IMA Journal of Management Mathematics* 4, 43–51 (1992)
11. Frydman, H.E., Altman, E.I., Kao, D.L.: Introducing recursive partitioning for financial classification: the case of financial distress. *Journal of Finance* 40(1), 269–291 (1985)
12. Dietterich, T.G.: Ensemble methods in machine learning. In: Kittler, J., Roli, F. (eds.) *MCS 2000. LNCS*, vol. 1857, pp. 1–15. Springer, Heidelberg (2000)
13. Breiman, L.: Random Forests. *Machine Learning* 45(1), 5–32 (2001)
14. wikipedia, http://en.wikipedia.org/wiki/Random_forest
15. Anita, P., Dirk, V.P.: Random Forests for multiclass classification: Random Multi Nomial Logit. *Expert Systems with Applications* 34, 1721–1732 (2008)
16. Verikas, A., Gelzinis, A., Bacauskiene, M.: Mining data with random forests: A survey and results of new tests. *Pattern Recognition* 44, 330–349 (2011)
17. Blake, C.L., Merz, C.J.: UCI repository of machine learning databases (1998)
18. Yang, Y.: An evaluation of statistical approaches to text categorization. *Information Retrieval* 1(1-2), 69–90 (1999)

A Study of Dynamic Decoupling Fuzzy Control Systems by Means of Generalized Inverse Matrix

Qing-Min Fan

Department of Mathematics, Taiyuan University of Technology
030024 Taiyuan, China
tyutfqm@163.com

Abstract. In the dynamic decoupling fuzzy control system, if singular or rectangle matrices are encountered in the state equation, the only way is to consider the decoupling control over partial states of the system. This paper presents a new conclusion with fewer parameters by virtue of generalized inverse matrix and realized decoupling fuzzy control over the system with multiple variables, which demonstrates good control effect.

Keywords: Nonlinear dynamic inverse, Decoupling control, Generalized inverse matrix, Fuzzy controller.

1 Introduction

In the study of nonlinear dynamic inverse, we can introduce nonlinear input to counteract the influence of the nonlinear factors in the system and obtain the expected linear dynamic model in the decoupling way. If the method of dynamic inverse and that of fuzzy control are combined, then the joint method will possess advantages of both. For convenience and without loss of generality, we take the linear multiple variable system into account, but the conclusions obtained can also be applicable in the nonlinear multiple variable system.

Let

$$\dot{x}(t) = Ax(t) + Bu(t) \tag{1}$$

be the state equation of the system.

If B can be invertible, then the decoupling control of the system can be realized. But in most circumstances, B is a rectangle matrix, and thus what we can do is to reduce (1) to a form of block, and then realize decoupling control over partial states of the system through dynamic inverse method [1~3]. In this paper, we employ the concept of generalized inverse of a singular or rectangle matrix to realize the decoupling fuzzy control over the system (1) with multiple variables.

2 Preliminaries

Let C^n be the complex n -dimension vector space and $C^{m \times n}$ be the set of $m \times n$ matrices and $C_r^{m \times n} = \{A \in C^{m \times n}, \text{rank } A=r\}$. Assume $R(A) = \{y \in C^m : y = Ax, x \in C^n\}$ is the range of A and $N(A) = \{x \in C^n : Ax=0\}$ is the null space of A . Let A^H denote the conjugate transpose matrix of a complex matrix A .

Definition 1[4]

Let $A \in C^{m \times n}$, if $X \in C^{n \times m}$ satisfies

$$\begin{aligned} AXA &= A, \\ XAX &= X, \\ (AX)^H &= AX, \\ (XA)^H &= XA. \end{aligned}$$

then X is called Moore-Penrose generalized inverse of A , denoted by A^+ .

The four equations in definition 1 are called Moore-Penrose equations. Each equation has some meaning and can be conveniently applied in practice[5,6]. For different purpose, one usually considers X that partially satisfies the equations, which is referred to as weak inverse.

Definition 2

Let $A \in C^{m \times n}$ and $A\{i, j, \dots, l\}$ denote the set of all matrices $X \in C^{n \times m}$ which satisfy the i -th, j -th, ... and l -th Moore-Penrose equation. A matrix $X \in A\{i, j, \dots, l\}$ is denoted by $X = A^{(i, j, \dots, l)}$ and is referred to as the $\{i, j, \dots, l\}$ -inverse of A .

In this paper, we present a new result expressing the set $A\{1\}$ ($A^{(1)} \in C^{n \times m}$), which involves the fewest arbitrary parameters. With the expression, the decoupling fuzzy control is more effective.

Lemma 1[7]

Let $A \in C^{m \times n}$, $b \in C^m$. Then the system of linear equations

$$Ax = b \tag{2}$$

is consistent if $AA^{(1)}b = b$ for all $A^{(1)}$.

In this case, the general solution of (2) is

$$x = A^{(1)}b + (I - A^{(1)}A)y$$

where $y \in C^n$ is arbitrary.

Lemma 2[7]

Let $A \in C_r^{m \times n}$, $r > 0$. Then there exists a column full-rank matrix $F \in C_r^{m \times r}$ and a row full-rank matrix $G \in C_r^{r \times n}$ such that $A = FG$.

Lemma 3[7]

If $A^{(1)} \in A\{1\}$, then $R(AA^{(1)}) = R(A)$, $N(A^{(1)}A) = N(A)$ and $R((A^{(1)}A)^H) = R(A^H)$.

3 Applying Generalized Inverse of Matrix to Realize Decoupling Fuzzy Control over the System

Theorem 1

Let $E \in C^{n \times n}$ be an idempotent matrix .Then

- 1) E^H and $I - E$ are idempotent;
- 2) The eigenvalues of E are 0 and 1, and the multiplicity of 1 is rank E ;
- 3) rank $E = \text{trace } E$;
- 4) $E(I - E) = (I - E)E = O$;
- 5) $Ex = x$, iff $x \in R(E)$;
- 6) $N(E) = R(I - E)$.

Proof

The proof of 1)~5) is straightforward by the definition of idempotency.

By 2) and the fact that the trace of every square matrix equals the sum of all its characteristic values as a repeated root is calculated as many times as its multiplicity, we can easily obtain 3).

6) is verified by applying Lemma 1 to the equation $Ex = 0$.

Theorem 2

Let the square matrix E has the full-rank decomposition $E = FG$. Then E is idempotent iff $GF = I$.

Proof

If $GF = I$, then we have

$$(FG)^2 = FGFG = FG. \tag{3}$$

Conversely, since F is a column full-rank matrix, its $\{1\}$ -inverse is its left inverse. Similarly, G is a row full-rank matrix implies that its $\{1\}$ -inverse is its right inverse. Hence

$$F^{(1)}F = GG^{(1)} = I.$$

Now It follows from (3) that $GF=I$ by multiplying $F^{(1)}$ on the left and $G^{(1)}$ on the right.

Theorem 3

Let $A \in C_r^{m \times n}$, $A^{(1)} \in A\{1\}$ be any fixed element and let $F \in C_{n-r}^{n \times (n-r)}$, $M^H \in C_{m-r}^{m \times (m-r)}$, $P \in C_r^{n \times r}$ be any given matrices whose columns are a basis of $N(A)$, $N(A^H)$ and $R(A^{(1)}A)$ respectively. Then the general solution of Moore-Penrose equation

$$AXA = A \tag{4}$$

is

$$X = A^{(1)} + FY + PZM \tag{5}$$

where $Y \in C^{(n-r) \times m}$, $Z \in C^{r \times (m-r)}$ are arbitrary, i.e.

$$A\{1\} = \{A^{(1)} + FY + PZM: Y \in C^{(n-r) \times m}, Z \in C^{r \times (m-r)}\}. \tag{6}$$

Proof

It is obvious that $AF = O$, $MA = O$, and thus the right side of (5) satisfies (4), i.e. $X \in A\{1\}$. It follows from Lemma 3 and 6) in Theorem 1 that

$$R(I_n - A^{(1)}A) = N(A) \text{ and } R((I_m - AA^{(1)})^H) = N(A^H).$$

By Lemma 2, there must exist unique matrices G , H , D such that $FG = I_n - A^{(1)}A$, $HM = I_m - AA^{(1)}$, $PD = A^{(1)}A$.

Since these multiplications are idempotent, by Theorem 2,

$$GF = I, \quad DP = I, \quad MH = I. \tag{7}$$

and

$$GP = O, \quad DF = O \tag{8}$$

are obviously valid.

From (5), (7) and (8), it follows that

$$Y = G(X - A^{(1)}), \quad Z = D(X - A^{(1)})H. \tag{9}$$

Now let X be an arbitrary element in $A\{1\}$. Substituting Y and Z in (9) into (5), we can see that (5) is satisfied. Therefore, (5) is the general solution of (4).

In the system (1), if B is rectangle matrix, let

$$u(t) = B^{(1)}[V(t) - Ax(t)],$$

where $B^{(1)} \in B\{1\}$, $V(t) = \dot{x}(t)$ is the state equation of the system. Choose $V(t)$ as

$$V(t) = K[x_c(t) - x(t)] \tag{10}$$

where K is a diagonal matrix and c stands for instruction.

By (10), we have

$$\dot{x}(t) = K[x_c(t) - x(t)] \tag{11}$$

Hence the decoupling control is realized in the system with multiple variables. By choosing the diagonal elements of the matrix K , the closed loop pole of the system can be arbitrarily set.

By means of self-tuning method of fuzzy control rule with correct factors, a well-behaved fuzzy controller can be designed by optimizing the correct factors, proportion and integral factor. The dynamic performance of the system output will be significantly improved by replacing the matrix K in (11) and adjusting fuzzy control rules and optimizing the parameters involved in the fuzzy controller.

References

- [1] Zhang, W., Lu, J., Wu, F.: Introduction to Advanced Control Theory and Method, pp. 98–162. Northwest University of Technology Press, Xian (2000) (in Chinese)
- [2] Cheng, P., Wang, Y.: Fundamentals of Modern Control Theory, pp. 1–143. Beijing Aerospace University Press, Beijing (2004) (in Chinese)
- [3] Gong, L.: Modern Tuning Techniques-Basic Theory and Analytic Approaches, pp. 133–291. South-east University Press, Nanjing (2003) (in Chinese)
- [4] Horn, R.A., Johnson, C.R.: Matrix Analysis, pp. 198–263. Cambridge University Press, Landon (1985)
- [5] Chen, Y.: Disturbance of Matrix and Its Generalized Inverse. Journal of Applied Mathematics 9(2), 319–327 (1986)
- [6] Xu, Z., Lu, Q.: Inverse and Generalized Inverse of Centrally and Anti-centrally Symmetric Matrix. Shuxue Tongbao 7, 37–42 (1998)
- [7] Jiang, Z., Shi, G.: Matrix Theory and Its Applications, pp. 87–192. Beijing Aviation College Press, Beijing (1988) (in Chinese)

Fractal Traffic Analysis and Applications in Industrial Control Ethernet Network

Sen-xin Zhou^{1,2}, Jiang-hong Han¹, and Hao Tang¹

¹ School of Computer and Information, Hefei University of Technology,
Hefei Anhui, 230009, China

² Management science and engineering School of Anhui University of finance & economics,
Bengbu Anhui, 233041, China

ahcdzsx@126.com, hanjh@hfut.edu.cn, htang@hfut.edu.cn

Abstract. It has become clear that the traditional Poisson model of data network traffic is insufficient for dimensioning and analyzing the performance of real-life networks. Fractal models are more appropriate for simulating the self-similar behavior of data traffic. To understand self-similarity on physical grounds in a realistic network environment is important when developing efficient and integrated network frameworks within which end-to-end QoS guarantees are fully supported. OPNET features the Raw Packet Generator (RPG) which contains several implementations of self-similar sources. This paper uses fractal analysis to characterize increasingly bursty industrial control network traffic. The goal is to develop a better understanding of the fractal nature of network traffic, which in turn will lead to more efficiency and better quality of services on industrial control network traffic. We present a comparison between the different RPG models in OPNET Modeler.

Keywords: Industrial control network, Network traffic, Raw packet generator (RPG), Optimized Network Engineering Tool (Opnet).

1 Introduction

Recently, the industrial network becomes an indispensable component among automated systems. Especially, as the systems are required to be more intelligent and flexible, the systems should have more sensors, actuators, and controllers, often referred to as field devices. In most cases, these field devices require some type of electrical connection because they are distributed over a certain area. As the number of devices in a system grows and the functions of the system need to be more intelligent, these devices need to exchange the rapidly increasing amount of data among them. Conventionally, these devices are connected with point-to-point or direct connections, where each pair of devices requires at least one electrical cable. This approach is not suitable any more for the system composed of many devices because the number of cables is proportional to the square of the number of devices. As an alternative to the point-to-point connections, many industrial networks have been adopted, which can accommodate various data with shared transmission medium. Because the industrial network has more advantages than the point-to-point connection such as reduction of wiring

and ease of maintenance, its application areas have grown to include various applications such as process automation system, automated manufacturing system, and automated material handling system [1,2].

Industrial control systems are highly distributed networks used for controlling operations in water distribution and treatment plants, electric power systems, oil and gas refineries, manufacturing facilities and chemical plants. Generally, an industrial complex comprises two distinct networks: a process control network (PCN) containing controllers, switches, actuators and low level control devices, and an enterprise network incorporating high level supervisory nodes and corporate computers. PCN includes supervisory control and data acquisition systems and distributed control systems. The main components of a PCN are the control server or master terminal unit, remote terminal units, intelligent electronic devices, programmable logic controllers, operator consoles or human machine interfaces, and data historians. A trusted network (TN) architecture uses the existing standards, protocols, and hardware devices to extend the concept of trust to the network architecture. TNs provide important security services such as user authentication, comprehensive network device admission control, enddevice health check, policybased access control and traffic filtering, automated remediation of noncompliant devices, and auditing. The Trusted Computing Group (TCG) has promulgated industry standards for TNs. Several commercial TN technologies have been developed, including Cisco TrustSec, Cisco leanAccess, and Microsoft Network Access Protection. Cisco NAC is interoperable with Microsoft NAP. Trusted Computing is an industry initiative headed by the Trusted Computing Group. The main output of the group is a set of specifications for a hardware chip (the Trusted Platform Module) and surrounding software infrastructure like the TCG Software Stack.

Industrial Ethernet is trending to be the principal infrastructure choice for mission-critical industrial automation and control applications. Industrial Ethernet is built on the same standards-based networking platform as enterprise Ethernet, which has long reigned as the universal network solution. Industrial Ethernet connects the office with the plant floor by utilizing a single cabling platform with Ethernet connectivity and IP addressing. This convergence of open, standards-based Ethernet communications provides all the advantages of secure, seamless interoperability among manufacturing enterprise networks — from corporate offices to the shop floor — and enables Internet and enterprise connectivity, anytime and anywhere. In addition to system integration and interoperability, there are a host of other benefits to be gained from implementing a complete, end-to-end Ethernet solution — from cabling and connectivity, to active components and associated hardware. Key business benefits include lower overall Total Cost of Ownership (TCO) and higher return on investment (ROI) resulting from real-time visibility and flexibility, reduced network maintenance and administration costs and labor, and greater physical and virtual network security. Industrial communications and control networks are expected to operate consistently and reliably under extreme conditions, such as electromagnetic interference (EMI), high operating temperatures, ambient outdoor temperatures, power/voltage fluctuations, machine vibration, mechanical hazards and more. Given these environmental risks, it is clear that the networked communications systems in extreme environments must be exceptionally rugged and durable. Any physical deterioration or electrical failure in key data transmission

components can lead to unreliable network performance and safety issues, and may ultimately lead to loss of critical data, costly downtime, or even catastrophic failure.

The networks, systems, software applications, and data of many enterprises and organizations form a critical foundation and essential structure for industrial network. Without a reliable and functional network, the network control system is not secure. There are three key components of control networks analysis are network architecture, network protocols, and network performance analysis. The goal of a control network is to provide a guaranteed quality of service such as deterministic time delays and maximum throughput for real-time control applications. These networks target various types of industrial automation and processing applications and are distinguished through static parameters such as data rate, message size, medium length, supported topology, number of nodes, and dynamic parameters such as MAC mechanism, message connection type, interoperability, and interchangeability. In general, data exchanged on an industrial network can be classified into two groups: real-time and non-real-time data. Non-real-time data do not have stringent time limits on their communication delays experienced during the data exchange. In contrast, real-time data have very strict time limits and the data's value is diminished greatly as the communication delay grows larger. This real-time data can be further divided into periodic and asynchronous data, depending on the periodic nature of the data generation. On many industrial networks, these data types are sharing a single network although they have different requirements on communication. That is, the non-real-time data need assurance of delivery without error and duplication, while the real-time data are concerned mostly on the time taken to reach the destination.[2,3].

This paper uses fractal analysis to characterize increasingly bursty industrial control Ethernet network. It examines how the fractal nature of network traffic is developed due to the workings of network protocols and applications. The goal is to develop a better understanding of the fractal nature of network traffic, which in turn will lead to more efficiency and better quality of services on industrial control Ethernet network. In this paper we will study and compare the performance of an Ethernet segment run with heavy-tail traffic and with exponential traffic. The performance parameters considered here are link utilization and end-to-end delay. Our works are as follows: creation of the network model in opnet 14.5, running the simulation with different ON-OFF models and plotting and comparing the results[4].

This paper is organized into six sections including this introduction. Section 2 gives introduction for Basic fractal traffic theory, and RPG model architecture in section 3. Section 4 introduces RPG model model attributes. Experiments are presented in Section 5. Finally, summary and conclusions are presented in Section 6.

2 Basic Fractal Traffic Theory

Traffic self-similarity has a root from two fractal processes: Fractional Brownian Motion (FBM) and Fractional Gaussian Noise (FGN). The characteristics of FBM and FGN are represented by a fractal spectrum with a Hurst parameter, which can be easily analyzed using wavelet transform. A self-similar phenomenon represents a process displaying structural similarities across a wide range of scales of a specific dimension. In other words, the reference structure is repeating itself over a wide range of scales

and the statistics of the process do not change. However, these properties do not hold indefinitely for real phenomena and at some point, this structure breaks down. Self-similarity can therefore be associated with “fractals” which are objects with unchanged appearances over different scales. A stochastic process is called fractal when a number of relevant statistics exhibit scaling with related scaling exponents. Since scaling leads mathematically to power-law relationships in the scaled quantities the conclusion is therefore that the traffic shows fractal properties when several estimated statistics exhibit power-law behaviour over a wide range of time scales [5]. A continuous-time stochastic process $X(t)$ is considered to be statistical self-similar with parameter $H(0.5 \leq H \leq 1.0)$ if, for any real positive “ a ”, the process $a^H x(at)$ has the same statistical properties as $x(t)$. The parameter H is known as the Hurst parameter, or the self-similarity meter, and it is a key measure of self-similarity. More precisely H is a measure of the persistence of a statistical phenomenon and it is the measure of the length of the long-range dependence of a stochastic process. A value of $H=0.5$ indicates the absence of long-range dependence. The closer H is to 1 the greater the degree of persistence or long-range dependence. As a practical and effective method to bring guaranteed QoS support to control networks, priority queueing has currently become a popular research topic. However, there are many open issues concerned with the quantitative network performance, such as expected queue length, overflow probability, packet loss rate and so on. Our research is motivated by these issues, which are of significant importance due to the fact that statistically bounded traffic oriented performance evaluations lead directly into system and network design criteria for future control networks. On the other hand, recent studies have shown that existing packet switching networks experience selfsimilarity [6]. It has also been shown that the self-similar or long-range dependent stochastic process can better model high-speed networks than Poisson or Poisson-based processes, due to the capability of being able to capture the fractional property of network traffic across a wide range of time scales[7].

Traffic models are important to network performance analysis. They are required to capture the statistical characteristics of real traffic efficiently and accurately. The development of traffic modeling relies on the advances of traffic analysis. One of the traffic models is derived from an On-Off queuing model. In this model, “On” represents a busy data transmission period and “Off” represents silence with no data transmission. Statistically, if “On-Off” periods are heavy-tailed, aggregating a large number of these processes will result in long-range dependence. On the other hand, self-similarity of a process is a property represented using its power spectrum. By exploiting the power spectrum, a number of traffic models have been proposed. In particular, the Haar wavelet has been used to analyze and to model network traffic in recent years[8,11].

With traffic statistics that mathematically exhibit fractal characteristics: self-similarity and long-range dependence. With these properties, data traffic shows high peak-to-average bandwidth ratios and causes data networks inefficient. These problems make it difficult to predict, quantify, and control data traffic, in contrast to the traditional Poisson-distributed traffic in telephone networks. In this thesis, two analytical methods are used to study fractal network traffic. They are second-order self-similarity analysis and multifractal analysis. Using a number of experiments, the following results towards characterizing and quantifying the network traffic processes

have been achieved: First, self-similarity is an adaptability of traffic in networks. Many factors are involved in creating this characteristic. A new view of this self-similar traffic structure is provided. This view is an improvement over the theory used in most current literature, which assumes that the traffic self-similarity is solely based on the heavy-tailed file-size distribution. Second, the scaling region for traffic self-similarity is divided into two timescale regimes: short-range dependence (SRD) and long-range dependence (LRD). Experimental results show that the network transmission delay (RTT time) separates the two scaling regions. This gives us a physical source of the periodicity in the observed traffic. Also, bandwidth, TCP window size, and packet size have impacts on SRD. The statistical heavy-tailedness (Pareto shape parameter) affects the structure of LRD. In addition, a formula to quantify traffic burstiness is derived from the self-similarity property. Furthermore, studies of fractal traffic with multifractal analysis have given more interesting and applicable results. (1) At large timescales, increasing bandwidth does not improve throughput (or network performance). The two factors affecting traffic throughput are network delay and TCP window size. On the other hand, more simultaneous connections smooth traffic, which could result in an improvement of network efficiency. (2) At small timescales, traffic burstiness varies. In order to improve network efficiency, we need to control bandwidth, TCP window size, and network delay to reduce traffic burstiness. There are the tradeoffs from each other, but the effect is nonlinear. To apply this prior knowledge from traffic analysis and to improve network efficiency, a notion of the efficient bandwidth, EB, is derived to represent the fractal concentration set. Above that bandwidth, traffic appears bursty and cannot be reduced by multiplexing. But, below it, traffic is congested. An important finding is that the relationship between the bandwidth and the transfer delay is nonlinear.

The essential property of fractals is scale-invariant, and this property is often described as self-similarity in the second order statistics. In the early 1990s, Leland and Willinger et al. first reported the self-similar nature of network traffic. Before that study, traffic engineering had been successful in adopting a Poisson traffic model in traditional telephone networks. But, the new paradigm with fractal traffic has revolutionized traffic modeling. Among a number of network traffic models, fractional auto-regressive integrated moving average (FARIMA) and $M/G/\infty$ are two famous examples. FARIMA captures both short-range and long-range dependence. It has been successful in video traffic modeling. However, for generic network traffic, FARIMA lacks connections with the network mechanisms, e.g., TCP and HTTP. It also has a high computation complexity, which makes it difficult to use in practice. On the other hand, the $M/G/\infty$ model is based on the aggregation of general ON/OFF processes. This model is directly related to the heavy-tailed property of traffic, e.g., network objects such as traffic connections and file sizes that are heavy-tailed. In $M/G/\infty$, it is suggested that aggregating a large number of heavy-tailed processes results in long-range dependence. In the more recent studies, wavelet theory has been very popular in traffic analysis. A representation of traffic in wavelet domain requires only a few parameters across the scaling regions, which in turn leads to the wavelet traffic models including Independent Wavelet Model (IWM) and Multifractal Wavelet Model (MWM).

3 RPG Model Architecture

There are three PRG models working on MAC layer and network layer in opnet The following diagrams illustrate the underlying architecture of the RPG workstation node models:

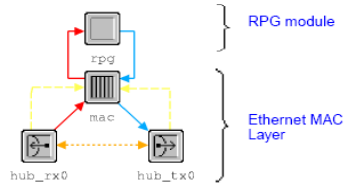


Fig. 1. Ethernet_rpg_station Node Model

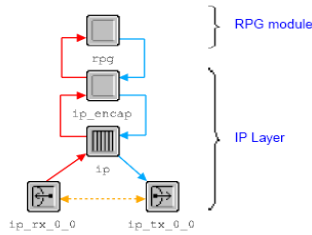


Fig. 2. ppp_rpg_workstation Node Model

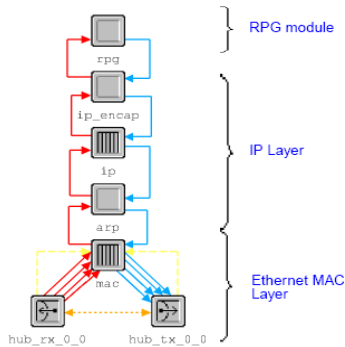


Fig. 3. Ethernet_rpg_wkstn Node Model

Each RPG node has an “rpg” module. The rpg module contains a dispatcher process that spawns a child process, called a manager process, for each arrival process used by the self-similar traffic source node. Each RPG node has an “rpg” module. The rpg module contains a dispatcher process that spawns a child process, called a manager process, for each arrival process used by the self-similar traffic source node. The

module.rpg_dispatcher process dispatches an RPG manager process for each row (arrival process) specified in the RPG Traffic Generation Parameters The table.rpg_mgr process spawns traffic source child processes. Each child process represents an independent traffic source. The type of child process generated (flow, ON-OFF, etc.) depends on which arrival process is used, while the number of child processes generated depends on the input parameters of that arrival process. The rpg_flow process is used by flow-based arrival processes to generate the packets within a flow. The rpg_onoff_source traffic source child process is used for all ON-OFF arrival The processes.rpg_source traffic source child process is used to generate packets in the Sup-FRP processes and to generate flows in the F-FSNDP processes[9,11].

4 RPG Model Model Attributes

All of the RPG nodes have an ‘‘RPG Traffic Generation Parameters’’ attribute that is used to specify the characteristics of the self-similar traffic. A source can generate self-similar traffic using one or more arrival processes. To use more than one arrival process, specify each process in a separate row of the RPG Traffic Generation Parameters table.

Average arrival rate: defines the average arrival rate for the aggregate traffic that is generated by all sources of the arrival process. The unit used for the arrival rate depends upon the particular arrival process being used. The Sup-FRP process and all three ON-OFF processes are defined in packets/sec, while the flow-based FSNDP processes are defined in flows/sec.

Hurst parameter: defines the Hurst characteristic of the self-similar traffic source. The Hurst characteristic determines the shape parameter for the Pareto distribution and is one of three main indicators of the traffic’s self-similarity. Fractal Onset Time Scale: the fractal onset time scale is used with the Hurst characteristic to determine the location parameter for the Pareto distribution. Like the Hurst characteristic, the fractal onset time scale is one of the main indicators of the traffic’s self-similarity.

Source activity ratio: defines the percentage of time that at least one of the independent ON-OFF traffic sources is active. This attribute is not used for other traffic sources (such as flow-based sources).

Peak-to-mean ratio: defines the ratio of peak traffic (that is, bursts) over the mean traffic rate, which is defined in the average arrival rate attribute. The peak-to-mean ratio attribute is used only with the ON-OFF arrival processes.

Average flow duration: defines the average duration of the flow for flow-based arrival processes. This attribute is not used in other,non-flow-based, arrival processes.**Filter Window Height:** defines the mean packet generation rate within a flow. This attribute is used only with flow-based arrival processes.

Input sup-FRP parameters: this compound attribute is used to specify the Hurst and fractal onset time scale parameters for the Sup-FRP process that is used as the flow arrival process by the F-FSNDP processes (that is, the shot-noise processes with fractal flow arrivals). The other arrival processes do not use this attribute.

Start time. This attribute is used to specify when the arrival process starts generating traffic. The default value of use global setting sets all of the arrival processes (for every node in the network) to begin traffic generation at the time specified in the RPG

start time simulation attribute. You can set a particular arrival process to start at a different time by specifying the desired time in its start time attribute. [10,11]

5 Experiments

OPNET is an Object Oriented (OO) discrete-event network simulator allowing for the modelling, implementation, simulation and performance analysis of communication networks and distributed applications. OPNET Modeller 14.5 is the industry's leading simulator specialised for network research and development. It allows to design and study communication networks, devices, protocols and applications with greater flexibility. It provides a graphical editor interface to build models for various network entities from physical layer modulator to application processes. All the components are modelled in an object-oriented approach which gives intuitive easy mapping to the real systems. It gives a flexible platform to test new ideas and solutions with low cost. OPNET is a simulator built on top of a discrete event system. It simulates the system behaviour by modelling each event happening in the system and processes it by user-defined processes. It uses a hierarchical strategy to organise all the models to build a whole network. OPNET also provides programming tools to define any type of packet format to be used in purpose-built protocols. Programming in OPNET includes the following major tasks: Define protocol packet format, define the state transition machine for processes running the protocol, define process modules and transceiver modules needed in each device node. It allows to finally define the network model by connecting the device nodes together using user-defined link models. We devised a set of Ethernet control network system model in opnet 14.5 simulation environment. We used RPG model, and different distribution for packet size (Pareto, exponential). From the simulations we see the differences between "old-fashioned" (exponential distribution) and "new fashion" (heavy-tailed distribution) methods of modeling self similarity network traffic. ExpON-PowOff model shows long periods with very small delays while exponential shows small variations and PowON-ExpOff to PowON-PowOff model shows somehow similar [11-15].

6 Conclusions

The main goal of this research work was to estimate the parameters of measured self similar traffic for modeling in OPNET. We use fractal analysis to characterize increasingly bursty industrial control network traffic. Our research is motivated by these issues, which are of significant importance due to the fact that statistically bounded traffic oriented performance evaluations lead directly into system and network design criteria for future control networks. Our goal is to develop a better understanding of the fractal nature of network traffic, which in turn will lead to more efficiency and better quality of services on industrial control network traffic. The future work is to develop trusted industrial control Ethernet network based on understanding of the self-similarity in industrial control Ethernet.

Acknowledgement. This paper is supported by National Nature Science Foundation of china under Project Number: 60473042;2009 Major Science Foundation Subject for the Education Department of Anhui Province under Project Number ZD200905 and 2010 Teaching & Research Foundation Subject the Education Department of Anhui Province under Project Number 20100473.

References

1. Park, K., Willinger, W.: Self similar network traffic and performance evaluation. Wiley, Chichester (2000)
2. Shao, Q.: Measurement and analysis of traffic in hybrid satellite-terrestrial network, Master of Applied Science Simon Fraser University (2004)
3. Gospodinov, M., Gospodinova, E.: The graphical methods for estimating hurst parameter of self similar network traffic. In: International Conference on Computing Systems and Tehnology – CompSysTech (2005)
4. Porwal, M.K., Yadav, A., Charhate, S.V.: Traffic Analysis of MPLS and Non MPLS Network including MPLS Signaling Protocols and Traffic distribution in OSPF and MPLS. In: International Conference on Emerging Trends in Engineering and Technology, ICETET (July 2008)
5. Ryu, B., Lowen, S.: Fractal Traffic Model of Internet Simulation. Network Analysis and Systems Dept. HRL Laboratories, of Saskatchewan (2002)
6. Gospodinov, M., Gospodinova, E.: The graphical methods for estimating hurst parameter of self similar network traffic. In: International Conference on Computing Systems and Tehnology CompSysTech (2005)
7. He, L., Botham, P.: Pure MPLS Technology. In: The Third International Conference on Availability, Reliability and Security. IEEE, Los Alamitos
8. Vonnahme, E., Ruping, S., Ruckert, U.: Measurements in switched Ethernet networks used for automation systems. In: Proceedings of 2000 IEEE International Workshop on Factory Communication Systems, pp. 231–238 (2000)
9. Porwal, M.K., Yadav, A., Charhate, S.V.: Traffic Analysis of MPLS and Non MPLS Network including MPLS Signaling Protocols and Traffic distribution in OSPF and MPLS. In: International Conference on Emerging Trends in Engineering and Technology, ICETET (July 2008)
10. Rahman, M. Kabir A.H., Lutfullah, K.A.M., Hassan, M.Z., Amin M.R.: Performance analysis of MPLS Protocols over conventional
11. RPG Model User Guide. OPNET Documentation
12. Configuring Applications and Profiles. OPNET Documentation
13. Simulation Methodology for Deployment of MPLS. OPNET Documentation
14. Chow, M.Y., Tipsuwan, Y.: Network-based control systems: a tutorial. In: 27th Annual Conference of the IEEE Industrial Electronics Society, Denver, pp. 1593–1602 (2001)
15. Walsh, G.C., Hong, Y.: Scheduling of networked control systems. IEEE Control Syst. Mag. 21, 57–65 (2001)

Routing on a Spherical Surface Using Hybrid PSO

Shoubao Su^{1,2}, Shuhao Yu¹, Yan Ma¹, Yang Yang¹, and Huali Xu¹

¹ Department of Computer Science & Technology, West Anhui University,
Lu'an 237012, Anhui, P.R. China

² Research Center of Satellite Technology, Harbin Institute of Technology,
Harbin 150080, Heilongjiang, P.R. China
{showbo, shyu, myan, xuhl, yangy}@wxc.edu.cn

Abstract. Routing on the surface of a sphere is a very interesting new topic. This paper presents a new discrete particle swarm algorithm, ENS-DPSO, to solve the travelling salesman problems on a spherical surface. Differently from previous approaches, ENS-DPSO redefines the path-relinking as velocity and position updating operators, and the hybridization with expanding neighborhood search (ENS) strategy is employed to improve the exploitation capabilities of the method. After visual implementation of the experimental tool in Java with 3D APIs, the effectiveness and efficiency of the proposed method are tested on various instances of random points with promising results.

Keywords: Particle swarm optimization (PSO), Traveling salesman problem (TSP), Routing on a sphere, Path-relinking, Expanding neighborhood search.

1 Introduction

Traveling Salesman Problem (TSP) is to find the shortest roundtrip of minimal total cost visiting each given node exactly once. More formally, a TSP instance is given by a fully connected graph G on a node set $V=\{1,2,\dots,m\}$, for some integer m , and by a cost function assigning a cost d_{ij} to the arc (i,j) , for any i, j in V . The TSP is one of the most widely known combinatorial optimization problems, since it is simple to describe and widely applicable but very difficult to solve, and actually, it is the NP-complete problem. Whereas, the extreme intractability of TSP has still attracted great interest from many researchers to attempted to develop new TSP-solving methods to find the optimal solution or quasi-optimal solution within a reasonable time for a given set of cities[1][2]. The harder is the problem researched on, the more significant is the result obtained. Since the limitation of traditional mathematical method to solve the large TSP instances, many intelligent methods based on evolutionary algorithms have been developed to yield good solutions within a acceptable computational time[2]. Particle swarm optimization (PSO) is a collaborative population based evolutionary computation method inspired from flocking of birds or schooling of fishes. Being simple in concept, easy in implementation and quick in convergence, PSO has gained wide applications in a variety of fields successfully[3][4][5]. Although the developments of PSO have originally concentrated on optimization in

continuous search spaces, some researchers proposed various discrete particle swarm algorithms for solving different combinational problems, including TSPs[1][3][5].

Recently, route searching problems on the surfaces of 3D shapes has been drawn much attention[6][7][8]. Uğur [8] introduced a simple and effective method for TSP on the surface of a sphere, called SphereTSP problem, and the method was test by using GA with 2-opt for various instances. Sphere is one of the most common ball-shaped objects, and it is the thermodynamically most stable shape with the lowest overall surface energy. Sphere has some unique features such as low friction, rotating around any axis, with the smallest surface area. So, TSP on a sphere has a wide range of utilizes in science and engineering.

Inspired from both Marinakis[5] and Uğur[8], this paper introduces a method for calculating the shortest path between any two points on the spherical surface. Then, a hybrid particle swarm optimization algorithm with expanding neighborhood search (ENS) strategy is presented to solve the travelling salesman problems on the surfaces of a sphere, in which path-relinking strategy is used to update velocities and positions of particles in the swarm. After visual implementation of the experimental tool in Java with 3D APIs, the effectiveness and the performance of the proposed method are test on many sets of random points on a unit sphere.

2 Description of the TSP on a Sphere

TSP on a sphere, called SphereTSP, is a special case of the TSP on 3D objects. In this problem, a salesman or particle must visit n cities having coordinates on the spherical surface exactly once, return to the starting city with the shortest distance (minimize the total travel cost).

2.1 Representation of a Point on a Spherical Surface

On a sphere, involving to several related terminologies, a *great circle* is the intersection a plane and a sphere where the plane also passes through the center of the sphere, and consequently divides it into two equal parts. Great circles through the two poles are called lines (or *meridians*) of *longitude*, and circles on the sphere that are parallel to the equator are lines of *latitude* as shown in Fig.1(a).

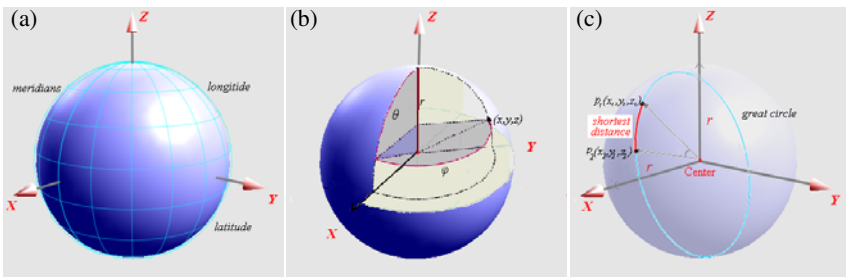


Fig. 1. (a) A sphere, (b) of coordinates, and (c) its shortest distance between two points on the spherical surface

The *shortest path* between two points on the spherical surface, called a *geodesic*, is along the segment of the great circle[9]. Every meridian of longitude is exactly half a great circle. The parallels of latitude are smaller circles except for the equator. A sphere all the points satisfying the following lie on a sphere of radius, r , centered at the origin $(0,0,0)$ as a function, $x^2+y^2+z^2=r^2$, of three space coordinates (x,y,z) .

The curvature of the curve of intersection is the sectional curvature. For the sphere each normal section through a given point will be a circle of the same radius, the radius of the sphere. It means that every point on the sphere will be an umbilical point. The Cartesian description for positions along the path of a curve can be parametrized via using the following vector point (r, θ, φ) transformation [9] such that

$$\begin{cases} x = r \sin \theta \cos \varphi \\ y = r \sin \theta \sin \varphi, 0 \leq \varphi \leq 2\pi \text{ and } 0 \leq \theta \leq \pi \\ z = r \cos \theta \end{cases} \quad (1)$$

2.2 Calculation of Shortest Distances Matrix

For any two points on the surface of a sphere there is a great circle through the two points. The minor arc of a great circle between two points is the shortest surface-path between them. The length of the minor arc of great circle is taken as the distance of two points on a surface of a sphere, namely *shortest distance*. The great circles are the geodesics of the sphere[9]. On a spherical surface, shortest distance (geodesics) between any pair of two points is along the arc of a great circle as shown in Fig.1(c). So, it can be used the value of angle θ in radians between two points. Given n points on a sphere, then the shortest arc length, d_{ij} , $i,j=1\dots n$, between two points (p_i, p_j) , can be calculated by using the following equation:

$$d_{ij} = ar \cos\left(\frac{x_i \cdot x_j + y_i \cdot y_j + z_i \cdot z_j}{r^2}\right), \quad i, j = 1 \dots, n \quad (2)$$

Since distance from any point p_i to point p_j is the same as the distance from point p_j to point p_i on the sphere, so, $n*n$ symmetric distance matrix $D=[d_{ij}]$ which gives the geodesics between all pairs of points on the spherical surface can be easily calculated by employing the Eq.(2).

3 Proposed Method

3.1 Particle Swarm Optimization (PSO) Algorithm

PSO is a swarm based evolutionary computation method and has applied to many fields in science and engineering[3][5][10]. It is initialized with a swarm of random solutions and this initial swarm evolves over generations to find optimum. In PSO, each particle i is represented by a position x_{ij} and a velocity v_{ij} , which enables them to fly through a d -dimensional problem space ($j=1\dots d$). The update of the position of a particle is performed by using its previous position information and its current velocity. Each particle knows its personal best position ($pbest_{ij}$) so far and the best

position achieved in the group (global best, $gbest$) among all personal bests. The basic equations for particle positions and velocities update are

$$\begin{cases} v_{ij}(t) = \omega v_{ij}(t-1) + c_1 r_1 (pbest_{ij}(t-1) - x_{ij}(t-1)) + c_2 r_2 (gbest_j(t-1) - x_{ij}(t-1)) \\ x_{ij}(t) = x_{ij}(t-1) + v_{ij}(t) \end{cases} \quad (3)$$

Where ω is inertia weight, acceleration factors c_1 and c_2 are nonnegative constants in the interval $[0, 1)$, and r_1 and r_2 are two random numbers drawn from $[0, 1)$. PSO algorithm updates the particles' positions and velocities iteratively with time step t until a stopping criterion is met.

3.2 Velocities and Positions Updating by Path Relinking

Path-relinking (PR) was originally suggested [5] as an approach to integrate intensification and diversification strategies in the context of Tabu search. This approach generates new solutions by exploring trajectories that connect high-quality solutions, by starting from an *starting particle*, and generating a path in the neighbourhood space that leads toward the *target particle* [11][12]. To generate the desired paths, the roles of starting and target particles can be interchangeable. Each particle can also be induced to move simultaneously toward the other as a way of generating combinations.

A particle in PSO can either follow its own way, or go back to its previous optimum, or go towards to the global optimum. Thus, in the hybrid method when the particles decides their next positions, a PR strategy is applied where the current particle plays the role of the starting particle and the best particle of the swarm or the current best particle plays the role of the target solution. The trajectories between the two particles are explored by simple swapping of the two points of the starting particle until the starting solution becomes equal to the target solution. If in some step of the PR strategy a new best solution, either of the particle or of the whole swarm, is found then the current best particle is replaced with the new one and the algorithm continues. An example of path-relinking operator is illustrated as shown in Fig.2.

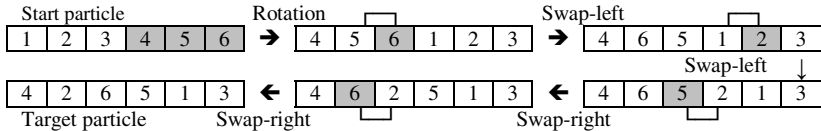


Fig. 2. An example of path-relinking operator

3.3 Expanding Neighborhood Search (ENS)

The hybridizations with local search strategies are often adopted by researchers in order to improve the exploration capability of many heuristic algorithms. On basis of CRLSM, Circle Restricted Local Search Moves strategy[12], Hansen[13] proposed a Expanding Neighborhood Search (ENS) strategy to use a larger neighborhood to escape from a local minimum to a better one. The method starts with a given size of neighborhood (the radius of the circle). Inside this neighborhood, a number of

different local search strategies are applied until all the possible trial moves have been explored and the solution can not further be improved in this neighborhood. Subsequently, the neighborhood is expanded by increasing its size by a percentage and, again, the same procedure is repeated until the stopping criterion is activated the algorithm has reached the maximum number of iterations, then the algorithm terminates with the current solution. An example of ENS strategy is illustrated in Fig.3 adopted from Ref[12], and its pseudocode is rewritten as following:

//Algorithm: ENS-Expanding Neighborhood Search

Initial size of neighborhood, $f = A/2$, the cost of the pending deletion edges, A, B

Do while ($\omega > e$ or $f < A + B$) // the quality of the solution, ω

$m = 1$ // index for the local search methods, m

Call first local search strategy

If $\omega < e$ then // a prespecified threshold number, e, b_1

STOP with the current solution

Else

Do while ($m < M_1$) // the number of local search strategy, M_1

If $\omega < b_1$ or a local optimum is found then

Change local search method

$m = m + 1$

If $\omega < e$ then

STOP with the current solution

Endif

Endif

Enddo

Endif

$f = f * (1+h)$ // the expanding rate of neighborhood, h , default 10%

Update b_1 and e

Enddo

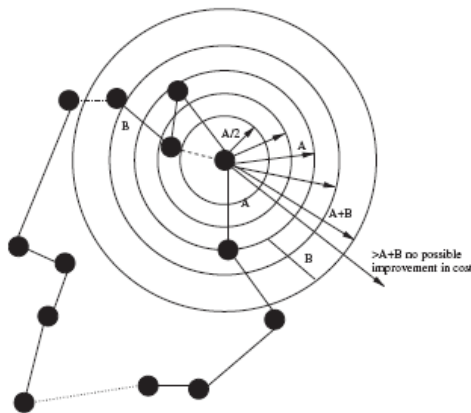


Fig. 3. An example of Expanding Neighborhood Search strategy[10]

3.4 Discrete PSO for Solving SphereTSP

Since the original PSO was applicable to continuous optimization problems, in last years several modifications of the algorithm known as Discrete PSOs have been proposed by many researchers and effectively applied with some discrete optimization problems. In this subsection, a new algorithm based on discrete PSO is proposed that incorporate extending neighborhood search strategy into and utilizes a feedback procedure in order to extend the exploration abilities of the discrete PSO.

The proposed algorithm initially defines the probabilities associated with each velocity for particles, where pr_1 , pr_2 , pr_3 correspond, respectively, to the likelihood that the particle follows its own way, goes toward the bet previous position and goes toward the best solution so far. Then, the algorithm proceeds modifying the particle's position according to the velocity operator randomly chosen. Finally, the probabilities are updated. At the beginning, a high probability is set to pr_1 , and low values are assigned to pr_2 and pr_3 . The goal is to allow that individualistic moves occur more frequently in the first iterations[1][3]. During the execution this situation is being modified and at the final iterations, pr_3 has the highest value. The idea is to intensify the search in good regions of the search space in the final iterations. Thus, the general framework of the algorithm for SphereTSP problem is described as follows:

```
//Algorithm: ENS-DPSO for SphereTSP
Setup initial parameters
Input the size of the swarm,  $m$ 
Generate randomly the initial particles
Define initial probabilities,  $pr_1$ ,  $pr_2$ ,  $pr_3$ 
Do While (the maximum iterations or a stop criterion is not met)
  For each particle  $p_i$ 
    Evaluate the fitness( $p_i$ )
    Update the personal optimum,  $pbest_i$ 
    Apply ENS
  Endfor
  Update the global solution,  $gbest$ 
  For each particle  $p_i$ 
    Update velocity of particle  $p_i$  using PR operators
    Update position of particle  $p_i$  using PR operators
  Endfor
  Update probabilities:
     $pr_1 = pr_1 * 0.95$ ,  $pr_2 = pr_2 * 1.01$ ,  $pr_3 = 100\% - (pr_1 + pr_2)$ 
Enddo
Return the best particle (the shortest tour)
```

4 Computational Experiments

To test the effectiveness and efficiency of the proposed algorithm, the visualization experimental tool for TSPs on a sphere benefited from Uğur's demo[8] reusable online, is designed in Java with 3D APIs, which are application programming

interfaces used for writing 3D graphics applications in geometric rendering and visually present all the sphere, points and route of the problems. The system provides some functions to allow user to interactively specify a new sphere, select a test data or add random points on a surface of the sphere, set parameters for the algorithm, and to look at the found optimal route in a 3D view as shown in Fig.4.

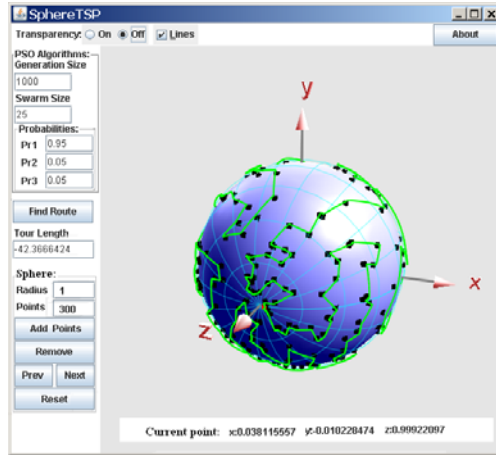
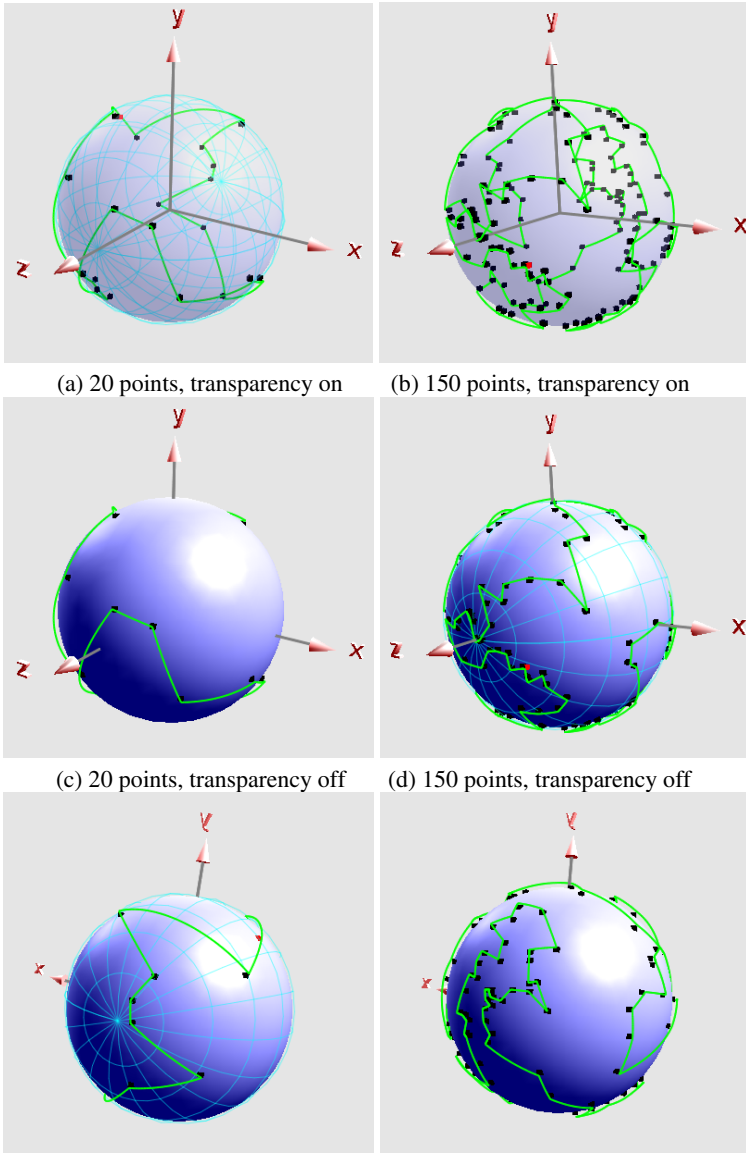


Fig. 4. Main interface of the experimental tools for SphereTSPs

In the experiment, initial parameters for ENS-DPSO algorithm are set to their default values as previous literature suggested, path-relinking with 3-opt exchanges selected, the initial probabilities $pr_1=0.95$, $pr_2=pr_3=0.05$. Several sets of random instances, e.g., $n = 10, 20, 30, 50, 100, 200, \dots, 800$ points, are respectively placed on a unit sphere. Each instance is run for 1000 iterations and independently 30 replications are conducted for each instance. Since the optimum solutions for random instances are unknown, the proposed method is able to find the best optimum for each one. The statistical average routes and standard deviation among 30 replications for each instance are recorded to evaluate the performance of the proposed algorithm. In replication trials for some low-dimension instances, such as $n < 300$ points, the optimum routes are frequently obtained from the algorithm. For larger scale problems, however, $n > 500$ points, the algorithm cost more computational time. As the density of points is increasing, the found best routes and the average routes are grow just as expected, while the average shortest arc length between two points decreases. Computational experiments have shown that the proposed algorithm is not only highly effective, but also can find closer sub-optimal route with high frequency for all the tested instances.

The experimental tool also allows users to visually understand the problem and examine the quality of the solution. The solid mode (transparency off) of the system permits examination of the route by interactive rotation of the sphere. For example, all cities and the optimal routes found by using ENS-PSO method for 20 points and 150 points are shown in Fig.5(a), (c), (b) and (d), and after rotating the tour on other of the spherical surfaces can be seen in Fig.5(e) and (f), respectively.



(a) 20 points, transparency on

(b) 150 points, transparency on

(c) 20 points, transparency off

(d) 150 points, transparency off

(e) 20 points, transparency off, after rotating (f) 150 points, transparency off, after rotating

Fig. 5. Tours found by ENS-DPSO for two random points on the surface of unit sphere

5 Conclusions

The research has scientific significance and practical value, since a sphere is one of the most common ball-shaped objects in nature. In this work a simple and efficient method to search the optimum route on a spherical surface has been proposed by

exploring the shortest arc of a great circle. Then, by employing path-relinking strategy as velocity and position updating operators, a new discrete PSO algorithm has been proposed by cooperating ENS local search strategy into the approach to solve the travelling salesman problems on the surfaces of a sphere. After implementation in Java with 3D APIs, various sets of random points are used to test the efficiency and the effectiveness of the proposed algorithm with satisfactory results.

The future study should investigate the comparison the proposed algorithm with other heuristic methods, and further develop an effective hybrid method base on swarm intelligence to solve the generalized TSP on a sphere in order to meet the practical needs of a wide range of engineering applications.

Acknowledgments. The work was supported in part by the National Natural Science Foundation of China under grant No.61075049, and the Projects of Natural Scientific Research Funds from Education Bureau of Anhui Province of China under grant Nos.KJ2011A268, KJ2010B471, KJ2011Z399, and KJ2010Z400.

References

1. Elizabeth, F.G., Marco, C.G., Givanaldo, R.: Particle Swarm Optimization Algorithm for the Traveling Salesman Problem. In: Greco, F. (ed.) *Traveling Salesman Problem*, pp. 85–96. Tech Publisher, Vienna (2008)
2. Tsai, H.K., Yang, J.M., Tsai, Y.F., et al.: An evolutionary algorithm for large travelling salesman problems. *IEEE Trans. Syst. CY B* 34, 1718–1729 (2004)
3. Mohemmed, A.W., Sahoo, N.C., Geok, T.K.: Solving shortest path problem using particle swarm optimization. *Applied Soft Computing* 8, 1643–1653 (2008)
4. Su, S.B., Cao, X.B., Kong, M.: Stability analysis of particle swarm optimization using swarm activity. *Control Theory & Applications* 27, 1411–1417 (2010)
5. Marinakis, Y., Marinaki, M.: A Particle Swarm Optimization Algorithm with Path Relinking for the Location Routing Problem. *J. Math. Model Algor.* 7, 59–78 (2008)
6. Ge, X.S., Bian, F.L.: On algorithm for 3D surface route optimization based on ant colony optimization. *Geomatics and Information Science of Wuhan University* 32, 366–372 (2007)
7. Takahashi, S., Fujimura, K., Tokutaka, H.: The SOM-TSP method for the three-dimension city location problem. *Neural Information Processing* 5, 2552–2555 (2002)
8. Uğur, A., Korukoğlu, S., Alişkan, A.Ç., et al.: Genetic algorithm based solution for TSP on a sphere. *Mathematical and Computational Applications* 14, 219–228 (2009)
9. Wikipedia.: Sphere from Wikipedia, the free encyclopedia Sphere (2011), <http://en.wikipedia.org/wiki/>
10. Anghinolfi, D., Paolucci, M.: A new discrete particle swarm optimization approach for the single-machine total weighted tardiness scheduling problem with sequence-dependent setup times. *Eur. J. Oper. Res.* 193, 73–85 (2009)
11. Glover, F., Laguna, M., Martí, R.: Scatter Search and Path Relinking: Foundations and Advanced Designs. In: Onwubolu, G.C., Babu, B.V. (eds.) *New Optimization Techniques in Engineering*, pp. 87–100. Springer, Berlin (2004)
12. Marinakis, Y., Marinaki, M., Dounias, G.: A hybrid particle swarm optimization algorithm for the vehicle routing problem. *Eng. Appl. Artif. Intel.* 23, 463–472 (2010)
13. Hansen, P., Mladenovic, N.: Variable neighborhood search: principles and applications. *Eur. J. Oper. Res.* 130, 449–467 (2001)

An Algorithm of Program Comprehension and Visual Representation for Object-Oriented Program

Hui Gu¹ and Daomiao Lin²

¹ Department of Computer, Zhejiang University of Technology,
310023 Hangzhou, China
gh@zjut.edu.cn

² Department of Computer, Zhejiang University of Technology,
310023 Hangzhou, China
lindaomiao@126.com

Abstract. An algorithm of object-oriented program comprehension and description is proposed in this paper. This methodology represents the program structure in visual form. Firstly, the code is analyzed. With the code analysis of compiling technique, knowledge related to the program structure can be concluded. Secondly, the knowledge is conducted for abstraction and inference. Lastly, software and algorithmic model are reconstructed. Program comprehension penetrates into algorithm structure analysis under function-level.

Keywords: Program comprehension, program analysis, information extraction, visual representation.

1 Introduction

Program comprehension is the procedure that analyses the code with computer technology automatically and gets the information about it [1][2]. It can review software documentation, such as software structure and algorithm description, with this information. Technical staff can understand the software effectively with the result of program comprehension.

There are some aspects need to be improved for existing tools of source program analysis, such as the speed, the output document stratified, algorithm description, context description [5][14][16]. An algorithm of object-oriented program comprehension and description is proposed in this paper. Basic approaches is: Firstly, the code is analyzed. With the code analysis of compiling technique, knowledge related to the program structure can be concluded. Secondly, the knowledge is conducted for abstraction and inference. Lastly, software and algorithmic model are reconstructed.[13]. These processes are shown in figure 1.

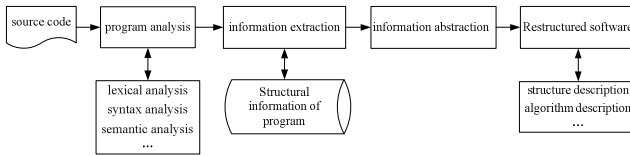


Fig. 1. Basic process of program comprehension

2 Regular Expression and Lexical Analysis

Regular expression is a special character string, composed by common characters and special characters (meta-characters). A series of character string construe regular expression and match target character string, according to specified grammar rules. This method can locate the target string quickly.

Meta-characters include: “*” matches 0 or multi-pattern; “+” matches 1 or multi-pattern; “?” matches 0 or 1 pattern; “.” matches any character except carriage return; “-” specifies range; “\” matches escaped meta characters; “[]” matches a character set. That matches any character of the parentheses. If the first character is “^”, it matches the negative mode [7].

Regular expression can match the target character string from code and get verbal messages at the lexical analysis. For example, regular expression of single-line comments is “//.*\n”; regular expression of multi-line comments is “/*(\s|.)?*\s*/”, “\s” matches the character of space, backspace, tabulator, linefeed and return. The code of the algorithm is as follow:

```

Algorithm NOTE(P)
Global: lid - index of P[0,...,m-1]
Input: P[0,...,m] - program array
Output: token - expression signature
begin
  C←P[lid++];
  while(C=Space∪C=Backspace∪C=Tab∪C=Enter∪C='/' )
  { if(C='/')
    { C←P[lid++];
      if(C='/')
      { while(C≠Enter){C•P[lid++];}
        token←signature of note;}
      else if(C='*')
      { notes←1;
        while(notes≠0)
        { C←P[lid++];
          if(C='*')
          { C←P[lid++];
            if(C='/'){notes=0;}else{lid--;}
          }
          token←signature of notes;}
        else{token←signature of '/';lid--;}
      }
    }
  }
  return token;
end
  
```

Lexical analysis matches the reserved word, special symbol and the identifier (variable, function and so on) accurately. Specific method is to design a lexical analyzer that is called by syntax analyzer. Lexical analyzer returns the lexical sign to syntax analysis procedure. At the beginning of lexical analysis, we must define a key vocabulary table to store key words of procedures language. Because there are various key words in source code, this table will be added dynamically. The steps of lexical analysis are:

- (1) Neglect space, carriage return, and other analogical characters;
- (2) Match words and special symbol;
- (3) Query the table and match key words;
- (4) Return lexical sign;

The structure of lexical sign [6]is:

```
struct Token{int syn;//expression code string text;//expression content};
```

3 Syntax Analysis and Semantic Analysis

Sub-analysis processes for each grammar rule are designed. The syntax analyzer chooses corresponding sub-analysis process, according to the lexical sign returned by the lexical analysis. In order to understand source code context, the syntax analyzer must be own semantic analysis, such as it have to review all the types in source code. According to c/c++ grammar, grammar rules of syntax analyzer are [8][9][10]:

```
<STMT>→<INCLUDE>|<CLASS>|<STRUCT>|<TYPEDEF>|...
```

The grammar of class definition is composed by key word, class name, inheritance and content.

```
<CLASS>→class<NAME><INHERIT><DEFINE>;
< INHERIT >→ε|:<PARENTS>
<PARENTS>→public<NAME>|<NAME>|<PARENTS>,public<NAME>
|<PARENTS>,<NAME>|...
```

Class members have their own access control permissions, default permission is private.

```
<DEFINE>→ε|{<MEMBER>}
<MEMBER>→ε|<MEMBER><ACCESS PERMISSION>
|<MEMBER><DATA MEMBER>|<MEMBER><MEMBER FUNCTION>
|<MEMBER><CLASS>|<MEMBER><STRUCT>|...
<ACCESS PERMISSION>→public|...
<DATA MEMBER>→<PREFIX><TYPE><NAME><VALUE>
<MEMBER FUNCTION>→
<PREFIX><TYPE><NAME>(<PARAMETER LIST>)<REALIZE>
<PARAMETER LIST>→
ε|<TYPE><NAME>|<PARAMETER LIST>,<TYPE><NAME>
<PREFIX>→ε|<PREFIX>const|<PREFIX>extern|<PREFIX>static|...
```

The grammar of function realization is composed by sequential control structure, loop control structure and branch control structure.

```

<REALIZE>→{<CONTROL STRUCTURE>}
<CONTROL STRUCTURE>→
  ε|<CONTROL STRUCTURE><SEQUENTIAL CONTROL STRUCTURE>
    |<CONTROL STRUCTURE><LOOP CONTROL STRUCTURE>
    |<CONTROL STRUCTURE>< BRANCH CONTROL STRUCTURE>
<SEQUENTIAL CONTROL STRUCTURE>→
  <EXPRESSION><CONTROL STRUCTURE>
<LOOP CONTROL STRUCTURE>→
  while(<EXPRESSION>){<CONTROL STRUCTURE>}
  lfor(<EXPRESSION>){<CONTROL STRUCTURE>}|...
<BRANCH CONTROL STRUCTURE>→
  if(<EXPRESSION>){<CONTROL STRUCTURE>}
  else{<CONTROL STRUCTURE>}
  lswitch(<EXPRESSION>){<CASE CONTROL STRUCTURE >}|...
<CASE CONTROL STRUCTURE >→
  ε|<CASE CONTROL STRUCTURE >
  case< EXPRESSION >:<CONTROL STRUCTURE>
    |<CASE CONTROL STRUCTURE >default:<CONTROL STRUCTURE>
<EXPRESSION>→<DEFINE>|<OPERATION>

```

Variable identifier is character string, whose first character must be a letter or underscore, and else characters must be composed by letter or figure.

```

<NAME>→<BEGIN><END>
<BEGIN>→A|...|Z|a|...|z|_
<END>→ε|<END>A|...|<END>Z|<END>a|...|<END>z|<END>_|<END>0|...
<TYPE>→void|int|string|vector|...|<TYPE>*<TYPE>&<TYPE>[|]...

```

Because the parsing is complex, here just list the part of grammar. The algorithm is as follow:

```

Algorithm PARSE(P)
  Input: P[0,...,m] - program array
  Output: null
  begin
    for i=0 to m-1
      { flag←True;
        token←SCANNER(P);
        while(flag=True)
          { if(syn∈#include){PARSER_INCLUDE(P);}
            else if(syn∈class){PARSER_CLASS(P);}
            else if(syn∈realization){PARSER_REALIZE(P);}...
          }
      }
  end

```

SCANNER is the lexical analyzer; PARSER_INCLUDE, PARSER_CLASS and PARSER_REALIZ are the sub-analysis processes of Syntax analyzer. Syntax analyzer chooses the sub-analysis processes, according to the value of syn. Such as algorithm of function realization analysis is:

```

Algorithm PARSE_REALIZE(P)
Input: P[0,...,m] - program array
Output: ControlStructure*
begin
  flag←True;
  token←SCANNER(P);
  while(flag=True)
  {  if(syn∈"if"∪syn∈"else"
      ∪syn∈"switch"∪syn∈"case"∪syn∈"default")
    {  BC•branch condition;SCSL•PARSE_REALIZE(P);
      add BC and SCSL into a controlStructure;
      add this controlStructure into the
      controlStructure list;}
    else if(syn∈"while"∪syn∈"for"∪syn∈"do-while")
    {  LC•loop condition;SCSL•PARSE_REALIZE(P);
      add LC and SCSL into a controlStructure;
      add this controlStructure into the
      controlStructure list;}
      ...
    else{ DE•expression;add DE into a controlStructure;
          add this controlStructure into the
          controlStructure list;}
  }
end

```

4 Information Description

Information description is divided into structure description and algorithm description. The description of structure adopts the type of class diagram; the description of algorithm adopts the type of N-S diagram. It needs to get the coordinate and size of the primitive graphics, the nesting and location relationship among themselves. It has to calculate the information(includes:coordinate and size)of its sub primitive graphics, before calculating the information of the primitive graphic. the algorithm of calculate the graphic size.

```

Algorithm CALCULATE_SIZE(C)
Input: C - ControlStructure
Output: size - the size of C
begin
  SC•the sub controlStructure list of C
  for i=0 to the size of SC-1
  {  P[i]=CALCULATE_SIZE(SC[i]);}
  return size•calculation(P);
end

```

5 Experimental Result of the Algorithm

According to above methods, we devise a system of program comprehension. the result of an instance is shown in figure 3 and figure 4.

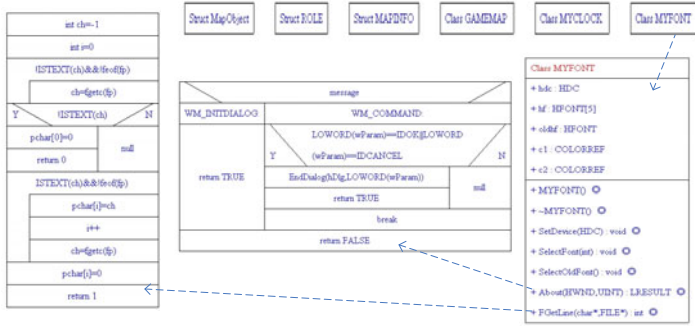
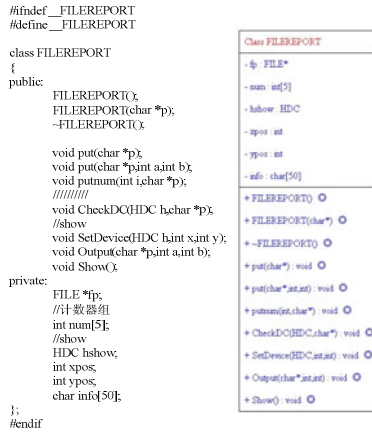
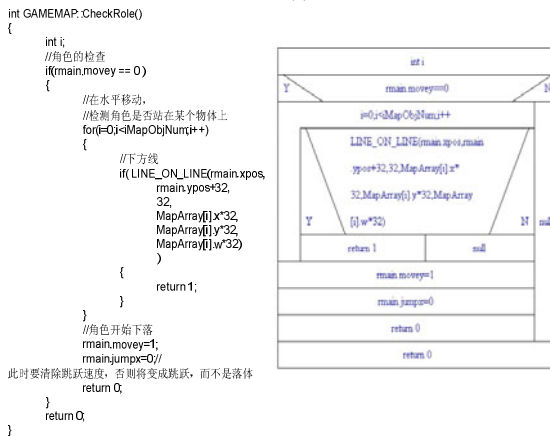


Fig. 3. A program comprehension instance



(a)



(b)

Fig. 4. Contrast between source code and result

Choose representative software from the open source, to test the prototype system of program comprehension. The result and reality of the software are in agreement. The test result is shown in table 1.

Table 1. Test result of prototype system

code number	Analysis time	Class (structure) information quantity				Function information quantity					
		data member		member function		Sequence Structure		loop structure		branch structure	
		code	accurate	code	accurate	code	accurate	code	accurate	code	accurate
5000	<5sec	174	171	175	175	1025	1023	53	53	144	144
1w	<10sec	354	354	1158	1158	3078	3078	80	80	507	507
5w	<5min	629	620	437	436	3798	3787	163	163	631	630
20w	<10min	2023	2014	7523	7521	23465	23459	496	496	2466	2466

References

1. Storey, M.-A.: Theories, Methods and Tools in Program Comprehension: Past, Present and Future. In: 13th International Workshop on Program Comprehension, pp. 181–191 (2005)
2. Grant, P.E.S., Chennamaneni, R., Reza, P.H.: Towards analyzing UML class diagram models to object-relational database systems transformations. In: DBA 2006 Proceedings of the 24th IASTED International Conference on Database and Applications, Anaheim, CA, USA, pp. 129–134 (2006)
3. Karahasanovic, A., Levine, A.K., Thomas, R.: Comprehension strategies and difficulties in maintaining object-oriented systems: An explorative study. *The Journal of Systems and Software* 80(9), 1541–1559 (2007)
4. Vinz, B.L., Etkorn, L.H.: Improving Program Comprehension By Combining Code Understanding With Comment Understanding. *Knowledge-Based Systems* 21(8), 813–825 (2008)
5. Understand. Source Code Analysis & Metrics [EB/OL] (November 20, 2010), <http://www.scitools.com/>
6. Levine, J.R., Mason, T., Brown, D.: *Lex and yacc*, pp. 43–151. China Machine Press, Beijing (2003)
7. Zhang, J., Zhang, Y.: Regular Expression and Application in Information Extraction. *Computer Knowledge and Technology* 5(15), 3867–3868 (2009)
8. Senlin, Y.: *Lemon Grammar Analysis of the Generator (LALR (1) type) Source Situation Analysis*, pp. 101–130. Zhejiang University Press, Hangzhou (2006)
9. Jiang, Z., Jiang, S.: *Formal Language and Automaton Theory*, pp. 42–82. Tsinghua University Press, Beijing
10. Sipser, M.: *Introduction to the Theory of Computation*, pp. 38–47. China Machine Press, Beijing (2000)
11. Kruglinski, D.J.: *Inside Visual C++, 4th edn.*, pp. 531–539. Tsinghua University Press, Beijing (2009)
12. Eckel, B., Allson, C.: *Thinking in C++, The second roll: Functional programming techniques*, pp. 124–130. China Machine Press, Beijing (2000)

13. Hui, G., Daomiao, L.: C/C++class relations visualization software: China0265128 [P] (January 12, 2011), <http://www.ccopyright.com>
14. Xing, Z., Jiahai, S., Fuqing, Y.: JBC+ + program understand tools. *Computer Engineering* 26(11), 80–81 (2000)
15. Hui, G., Huiying, Q.: Research and design of program abstract table based on UML class diagram. In: 2nd International Conference on Information Science and Engineering, ICISE 2010 (2010)
16. Imagix-4D. Comprehensive program understanding tool for C/C++ (October 1, 2010), <http://www.imagix.com/>

The Application of Simulated Algorithm Based on .NET in NP

Ailian Wang and Yuexing Duan

Department of Computer Education
Taiyuan University of Technology,
Taiyuan, Shanxi, 030024, P.R. China
ym4008cn@yahoo.com.cn

Abstract. Based on the principle of simulated annealing algorithm and applying annealing model, this paper systematically analyzed the extended applied algorithm and its examples of Nondeterministic polynomial (NP) decision problem, evaluated the setting of the important parameters and the processes, and realized the global optimum solution for NP application. The results of the experiments tested under .NET confirm that SA algorithm can jump out from the trap of local optimum and find global optimum solution for combinatorial optimization. Meanwhile, optimization algorithm is competitive comparing with other local searching algorithm.

Keywords: Simulated Annealing(SA), Metropolis Arithmetic, Nondeterministic Polynomial(NP), Cooling Schedule Formatting.

1 Introduction

SA algorithm was proposed at the earliest by Metropolis etc. (1953), and successfully applied to combinatorial optimization problem in 1983 by S.Kirkpatrick, C.D.Gelatt and M.P.Vecchi from IBM Thomas J. Watson Research Centre.[1][2]It is especially effective for complete combinatorial optimization problem. It is derived from a physical phenomenon of solid annealing process. After metal object is heated to a certain temperature, it is cooled down at specific speed. Slow cooling makes most of metal atoms reach the lowest energy point.. This can't be realized by means of quenching. Based on the similarity between annealing process and general combinatorial optimization problem, Local optimal solution can probably avoid being trapped in local minimum and get global optimum.

The rest of this paper is organized as follows. In section 2, the principle of SA algorithm and related theories are given. The procedure analysis appears in Section 3. In section 4, an example of extended SA algorithm was tested under .NET environment. Key parameters and operations of the algorithm are described and

researched in details in Section 5. NP realized in Section 6. Finally, the conclusions are given in Section 7.

2 Principle of SA Algorithm

2.1 Background of SA Algorithm

SA is originated from metallurgical terminology of annealing. Annealing is aimed to increase the size of crystal particle and reduce the defects existing in crystal lattice by means of specific cooling speed after the material being heated. The atom of the material stayed before at the position with local minimum internal energy. Temperature rising leads to the atom leaving the position it was with increased energy and moving randomly. The slow cooling speed by annealing can get crystal structure with low energy and minimize the probability of defective lattice[3].

2.2 Metropolis Criterion

Metropolis etc. proposed Importance Sampling in 1953, i.e., accepting new state by probability [4]. Specifically speaking, at temperature t , current state i produce a new state j , the energy for each state is E_i and E_j respectively. If $E_j < E_i$, then accepting new state as current; otherwise, if probability

$$p_r = \exp[-(E_j - E_i)/kt]. \quad (1)$$

is greater than the random number between $[0,1]$, still accepting new state as current. If equation is not, then keep i as current, k is Boltzmann constant.

This kind of sampling can accept the new state having great energy difference from the current state at high temperature, but just accept the new state with small energy difference from the current state at low temperature. And when the temperature tends to zero, it can't accept new state with higher energy than the current [5]. This accepted criterion is generally called Metropolis criterion. (The system will tend to a steady distribution state after a certain times of iteration).

2.3 NP Decision

For a decision problem, if there exist a polynomial function $g(x)$ and a verification algorithm A meeting with that it is the sufficient and necessary condition for making I is the 'Yes' example of the decision problem, i.e., there exist a character string 's' expressed as the feasible solution of I , the scale of it is $l(s)=O(g(l(I)))$, among of them, $l(I)$ is the scale of I , and the calculation time for that algorithm A verified that S is 'yes' answer of example I , then this decision problem is called nondeterministic polynomial, abbreviated as NP.

2.4 TSP Corresponding Decision Problem

Using once permutation of n cities (i_1, i_2, \dots, i_n) to express a business man departed from city i_1 and passed sequentially by i_2, \dots, i_n , and returned to i_1 again, according to such a path, decision problem is: for given z, if there exist a permutation of n cities

$$W = (i_1, i_2, \dots, i_n) \tag{2}$$

Which makes

$$f(w) = \sum_{j=1}^n d_{i_j i_{j+1}} \leq z. \tag{3}$$

Among, $i_{n+1} = i_1$.

One of the permutations meeting with the equation is called feasible solution of corresponding decision problem. The calculation time of verification algorithm is $n+1$, and the scale for the example l(I) is at least $2n(n-1)+2+\log_2 |p|$, Among,

$$p = n \prod \{d_y \mid d_y \neq 0, i \neq j\}. \tag{4}$$

Is the TSP example for any one city of n cities.

Its solution is $W = (i_1, i_2, \dots, i_n)$, a permutation of $(1, 2, \dots, n)$, the scale of the character string does not exceed

$$\sum_{i=1}^n (\lceil \log_2(i+1) \rceil) + 2n < 3n + n \log_2 n. \tag{5}$$

It can be decided that TSP belong to NP by verifying calculation time of the algorithm and judging the scale of feasible solution of the example.

2.5 Principle of SA Algorithm

SA algorithm, based on the similarity between practical annealing process and solving process for decision problem, corresponding the solution and objective function to the micro state and its energy, departing from a certain initial solution, by using of amount of transformation, finally find the relative optimum solution of combinatorial optimization for given control parameters. Furthermore, with the decreasing of the control parameters, by repeating execution of Metropolis algorithm, the global optimum solution can be obtained. SA algorithm is a general optimization algorithm, and it has been currently applied in engineering extensively [6].

3 SA Algorithm Procedure Analysis

Standard SA algorithm procedure can be described as the following:

Step 1. For given initial temperature, produce a initial state randomly, let $k=0$

Step 2. Repeat:

① Repeat produce a new state; $s_j = \text{Genete}(s)$; If,

$$\min\{1, \exp[-(C(s_j) - C(s))]\} \geq \text{random}[0,1]. \quad (6)$$

$s = s_j$. Until meeting with sampling stability criterion.

② Annealing $t_{k+1} = \text{update}(t_k)$, let $k = k + 1$, until meeting with algorithm ending criterion.

Step 3. Output result.

4 TSP Annealing Model

To TSP problem, the main principle of SA is to adopt random walking in neighborhood structure (random selecting points), by use of Metropolis sampling criterion, to make random walking gradually converge to local optimum solution. That is to iterate current solution to repeat process of produce new solution—calculating objective function difference---accept or discard, with the gradual attenuation of value t , to find the current solution as approximate optimum solution when algorithm ending. This is a heuristic random searching process based on Monte Carlo iteration.

Solution space: Solution space W is the all path of visiting each city thoroughly exact one time. Solution can be expressed as $\{ (i_1, i_2, \dots, i_n) \}$, i_1, i_2, \dots, i_n is a permutation of $1, 2, \dots, n$, representing departing from city i_1 , passing by cities i_2, \dots , in sequentially, and returning to city i_1 . Initial solution can be selected as $(1, \dots, n)$.

Objective function: objective function is the total length of the paths for visiting all cities (cost function);

$$f(i_1, i_2, i_3) = \sum_{j=1}^n (d_j, d_{j+1}). \quad (7)$$

Required optimum path is the corresponding path when objective function getting the minimum value.

Cooling parameters, region structure and new solution generator, accepted criterion and random number generator (Metropolis algorithm) constitute the three support of the algorithm. Cooling schedule should prescribe the following parameters:

- (a) . initial value t_0 of controlling parameter t
- (b) . attenuation function of controlling parameter t
- (c) . length of markov chain(i.e. for every random walking, the times of iteration after which it could tend to a quasi equilibrium distribution, i.e., a local convergence solution)
- (d) . selection for ending conditions

```

//cooling parameter
#define TIMES 500 //Searching times in solution space
at the same temperature
#define T 10000 //initial temperature
#define DecayScale 0.95 // Temperature Attenuation
factor (f(t)=t*DecayScale)
//Obtain a new solution, i.e., a new route.
void GetNewRoute(int*pos)
{
    int x1=0; int x2=0;
    for(;x1==x2;)
    {
        x1=rand()%Count; x2=rand()%Count; }
        int temp=pos[x1];
        pos[x1]=pos[x2]; pos[x2]=temp; }
// accept new solution or not
Bool IsAcceptNewRoute(double newResult,double
oldResult,double t)
{
    double _r=newResult-oldResult;
    double p=rand()%100;
    double pp=(p/100.00);
    if(t<=0.000000000000001)
        t+=0.00001;
    if(_r<0 || pp<exp(-_r/(t)))
    { return true; }
    else
    { return false; } }

```

5 Setting of Key Parameters and Operations

SA algorithm includes 3 functions and 2 criterions, i.e., state generation function, state accepted function, temperature update function, inner cycle ending criterion and outer cycle ending criterion. The design of these links will decide the performance of SA algorithm. In addition, initial temperature selection has a great influence to the algorithm.

State generation function: designing for state generation function (neighborhood region function) is to ensure that produced candidate solution distribute in whole solution space[7], which is commonly random variables generated by a probability distribution function.

State accepted function: commonly given by way of probability. Designing state accepted function obey the following principles.

At fixed temperature, the probability accepting candidate solutions which make the objective function decrease should be greater than the ones which increase objective function.

With the decrease of the temperature, the probability for accepting solutions which make objective function increase should decrease gradually. With the temperature tending to zero, only the solution which make objective function decrease can be accepted.

Test shows that the higher initial temperature, the higher the probability obtaining high quality solution. But calculation time will increase. Therefore, decision of initial temperature depend on test result, and should be adjusted certain times. Compromise should be considered between optimization quality and efficiency.

Temperature update function: used in outer cycle to modify temperature value. At present, in the theory of inhomogeneous SA algorithm convergency, temperature update can adopt function $t_k = \alpha / \log(k + k_0)$, but the temperature descends very slowly. In rapid SA algorithm, update function is $t_k = \beta / (1 + k)$, but the drop of temperature must match with state generation function as to ensure getting global optimum under reliable convergence. Generally, the more the candidate solution produced under different temperatures, the faster the temperature descends [8].

Inner cycle ending criterion, or Metropolis sampling stability criterion is used to decide producing candidate solution (state) numbers at different temperature. In the theory of inhomogeneous SA algorithm, due to only one or few amount of candidate solution being produced at one temperature, there does not exist the problem of selecting inner cycle ending criterion. But in the theory of homogeneous SA algorithm, convergence condition requires that candidate solution numbers produced at every temperature tend to infinite to make corresponding markov chain reach steady probability distribution. Obviously, it is impossible to be realized in practical algorithm application. Common-used sampling criterion include: (a). verify if average value of objective function is steady. (b). change of objective value is small for continuous calculation steps. (c). to sample at certain steps[9]. In addition, it can be adopted to use refuse times as a index to get the rule of controlling iteration steps. This article applied refuse times.

Outer cycle ending criterion: used to decide when algorithm ends. Setting final value of the temperature is a method[10]. If final value is reached and optimal value of objective function keeps unchanged at certain continuous steps, then optimizing ends.

Due to the control function of cooling schedule, the time complexity for SA algorithm is $O(kLmt(n))$, k is iteration times, Lm is the longest one in k Marcov chains, $t(n)$ is the polynomial function of problem scale n , which shows that reasonable selection for cooling schedule is the premise of algorithm application.

6 Example and Analysis

SA algorithm optimizer for TSP was tested under .NET environment. City node is stored in input file, output is the result. The optimum path obtained is showed in Fig. 1.

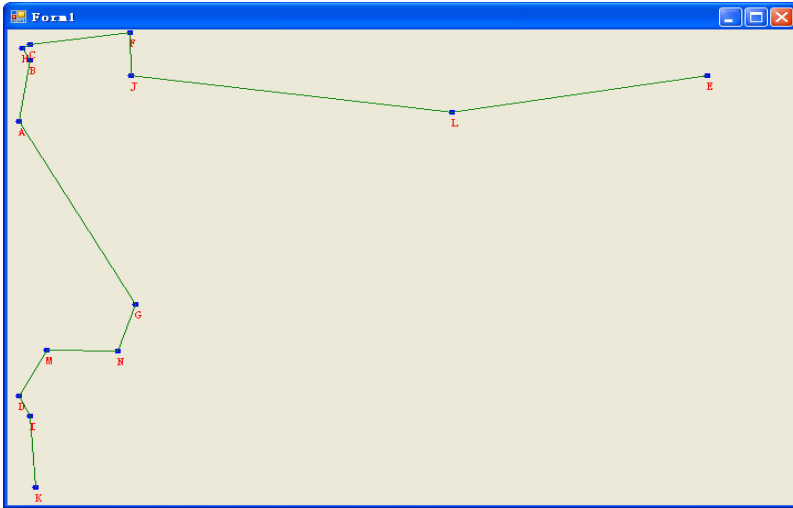


Fig. 2. Example of a extended SA algorithm

7 Conclusion

The feature of SA algorithm, which can avoid being trapped in local minimum, is significant for dealing with NP optimization for there existing more local minimum. Having more space, accepting new solution according to Metropolis criterion, not only accepting optimized solution, but also accepting worsen solution within some limitation [11], all these are nature difference between SA algorithm and local searching algorithm. SA algorithm can jump out from the trap of local optimum by change of temperature parameter in algorithm, and furthermore, can find global optimum solution for combinatorial optimization. It has the characteristics of simplicity and generality. From the view of time complexity, the average time complexity of local searching algorithm use polynomial of problem scale n as boundary, which is unknown under the worst condition. So, to most combinatorial optimization problems, SA algorithm is superior to local searching algorithm. But there are still some parameters needed to be adjusted, some rules to be matched. It needs large amount of numerical simulation calculation to be executed in order to select better parameter matching among the results.

References

1. Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E.: Equation of state calculations by fast computing machines. *J. Chem. Phys.* 21, 1087–1090 (1953)
2. Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P.: Optimization by simulated annealing. *Science* 220, 671–680 (1983)
3. Soleymani, S., Ranjbar, A.M., Shirani, A.R., Marami, M.: Application of Simulated Annealing Method in Strategic Bidding of Gencos. In: *The 21th International Power System Conference (PSC 2006)*, Tehran, Iran, pp. 1588–1594 (2006)

4. Zhang, X., Lin, X., Wan, C., Li, T.: Genetic-Annealing Algorithm for 3D Off-lattice Protein Folding Model. In: Washio, T., Zhou, Z.-H., Huang, J.Z., Hu, X., Li, J., Xie, C., He, J., Zou, D., Li, K.-C., Freire, M.M. (eds.) PAKDD 2007. LNCS (LNAI), vol. 4819, pp. 186–193. Springer, Heidelberg (2007)
5. Cheng, J.-K., Lee, H.-Y., Huang, H.-P., Yu, C.-C.: Optimal steady-state design of reactive distillation processes using simulated annealing. *Journal of the Taiwan Institute of Chemical Engineer* 40, 188–196 (2009)
6. Yu, C.-R., Luo, Y.: An Improved Nested Partitions Algorithm Based on Simulated Annealing in Complex Decision Problem Optimization. *WSEAS Transactions on Computers* 7, 75–82 (2008)
7. Saruhan, H.: Designing Optimum oil thickness in Artificial Human Knee Joint by Simulated Annealing. *Mathematical and Computational Applications* 14(2), 109–117 (2009)
8. Martins, T.C., Tsuzuki, M.S.G.: Rotational placement of irregular polygons over containers with fixed dimensions using simulated annealing and no-fit polygons. *Journal of the Brazilian Society of Mechanical Sciences and Engineering* 30(3), 205–212 (2008)
9. Cai, W., Ma, L.: Applications of critical temperature in minimizing functions of continuous variables with simulated annealing algorithm. *Computer Physics Communications* 181, 11–16 (2010)
10. Zhang, Y., Zhang, K., Zhou, Y., Kang, Y.: Common Mode EMI Suppression Based on Simulated Annealing Algorithm. In: WCICA 2006, Sixth World Congress on Intelligent Control and Automation, China, pp. 8448–8452 (2006)
11. Kolahan, F., Heidari, M.: Modeling and optimization of MAG welding for gas pipelines using regression analysis and simulated annealing algorithm. *Journal of Scientific & Industrial Research* 69, 177–183 (2010)

Coevolutionary Optimization Algorithm: With Ecological Competition Model

Jianguo Liu and Weiping Wu

Congqing Business and Technology University ,
400067 Chongqing, China
ljg65@163.com, wuweiping@ctbu.edu.cn

Abstract. Premature convergence and low converging speed are the distinct weaknesses of the genetic algorithms. a new algorithm called ECCA (ecological competition coevolutionary algorithm) is proposed for multiobjective optimization problems, in which the competition is considered to be in important position. In the algorithms, each objective corresponds to a population. At each generation, these populations compete among themselves. An ecological population density competition equation is used for reference to describe the relation between multiple objectives and to direct the adjustment over the relation at individual and population levels. The proposed approach store the Pareto optimal point obtained along the evolutionary process into external set, enforcing a more uniform distribution of such vectors along the Pareto front. The experiment results show the high efficiency of the improved Genetic Algorithms based on this model in solving premature convergence and accelerating the convergence.

Keywords: Competitive coevolutionary genetic algorithm, Pareto optimal point, multiobjective optimization problems.

1 Introduction

This solution of a real problem involved in multiobjective optimization (MO) must satisfy all optimization objectives simultaneously, and in general the solution is a set of indeterminacy points. The task of MO is to estimate the distribution of this solution set, then to find the satisfying solutions in it. General MO contain a set of n decision variables, a set of k objective functions, and a set of m constraints. In this case, objective functions and constraints respectively become functions of the decision variables. If the goal of multiobjective optimization problems is to maximize the objective functions of the y vector, then (1):

$$\begin{aligned} & \text{maximize } y = f(x) = (f_1(x), \dots, f_k(x)), \\ & \text{subject } e(x) = (e_1(x), \dots, f_j(x), \dots, f_m(x)) \leq 0 \end{aligned} \quad (1)$$

Where $x = (x_1, x_2, \dots, x_n) \in X$, $y = (y_1, y_2, \dots, y_k) \in Y$.

In (1), x is called a decision variable vector and y is called an objective function vector. The decision variable space is denoted by X and the objective function space is

denoted by Y . The constraint condition $e(x) \leq 0$ determines the set of feasible solutions. The set of solutions of multiobjective optimization problems consist of all decision vectors for which the corresponding objective vectors cannot be improved in any dimension without degradation in another [1]. Differently from Single-objective Optimization Problems (SOPs), multiobjective optimization problems have a set of solutions known as the Pareto optimal set. This solution set is generally called non-dominated solutions and is optimal in the sense that no other solutions are superior to them in the search space when all objectives are considered.

The set of objectives forms a space where points in the space represent individual solutions. The goal of course is to find the best or optimal solutions to the optimization problem at hand. Pareto optimality defines how to determine the set of optimal solutions. A solution is Pareto-optimal if no other solution can improve one objective function without a simultaneous deterioration of at least one of the other objectives. A set of such solutions is called the Pareto-optimal front.

Evolutionary algorithms (EA's) seem to be particularly suited for this task because they process a set of solutions in parallel, possibly exploiting similarities of solutions by recombination. Some researchers suggest that multiobjective search and optimization might be a problem area where EA's do better than other blind search strategies [2].

First, they can be applied when one has only limited knowledge about the problem being solved. Second, evolutionary computation is less susceptible to becoming trapped by local optima. This is because evolutionary algorithms maintain a population of alternative solutions and strike a balance between exploiting regions of the search space that have previously produced fit individuals and continuing to explore uncharted territory. Third, evolutionary computation can be applied in the context of noisy or non-stationary objective functions.

At the same time, difficulties can and do arise in applying the traditional computational models of evolution to multiobjective optimization problem. There are two primary reasons traditional evolutionary algorithms have difficulties with these types of problems. First, the population of individuals evolved by these algorithms has a strong tendency to converge because an increasing number of trials are allocated to observed regions of the solution space with above average fitness. This is a major disadvantage when solving multimodal function optimization problems where the solution needs to provide more information than the location of a single peak or valley [3]. This strong convergence property also precludes the long-term preservation of coadapted subcomponents required for solving covering problems or utilizing the divide-and-conquer strategy, because any but the strongest individual will ultimately be eliminated. Second, individuals evolved by traditional evolutionary algorithms typically represent complete solutions and are evaluated in isolation. Since interactions between population members are not modeled, even if population diversity were somehow preserved, the evolutionary model would have to be extended to enable coadaptive behavior to emerge [4].

To avoid this phenomenon, we proposed a multiobjective coevolutionary genetic algorithm (ECCA) for multiobjective optimization. Individual evolution is based on its fitness in genetic algorithm, but its living environment and relationship with other part aren't envolved. Coevolution is the process of mutual adaptation of two or more populations. The computational study of coevolution initiated by Hillis gave birth to competitive coevolutionary algorithms. His main motivation of the work reported

here was precisely to take advantage of some coevolutionary concept [5]. In 1994, Paredis introduced Coevolutionary Genetic Algorithms (CGAs) [6]. In contrast with the typical all-at-once fitness evaluation of Genetic Algorithms (GAs), CGAs employ a partial but continuous fitness evaluation. Furthermore, the power of CGAs was demonstrated on various applications such as classification, process control, and constraint satisfaction. In addition to this, a number of symbiotic applications have been developed.

The major concern of this paper is to introduce the idea of competitive coevolution into multiobjective optimization. A multiobjective competitive evolutionary genetic algorithm is proposed and implemented based on this idea to search the Pareto front effectively and reduce the runtime. At each generation, the proposed approach store the Pareto optimal point obtained along the evolutionary process into external set, enforcing a more uniform distribution of such vectors along the Pareto front. We then describe multiobjective optimization problems and a typical test functions that we use to judge the performance of the ECCA. Empirical results from the ECCA runs are presented and compared to previously published results.

The paper is organized as follows. Section 2 gives the ecological population competition mode and the proposed algorithm. In Section 3, he results of simulation are presented. The conclusions are given in Section 4.

2 The Proposed Algorithm

In collaborative evolution individual's self-status, living environment and competition with other individuals affect individual's self-evolution [7]. Lotka-Volterra competition equation as the theoretical model of population competition is introduced to describe populations' cooperation. Given two populations N_1 , N_2 , the cooperation between them can be formulated as follows (2) and (3):

$$\frac{dN_1}{dt} = r_1 N_1 \left(\frac{K_1 - N_1 - a_{12} N_2}{K_1} \right) \quad (2)$$

$$\frac{dN_2}{dt} = r_2 N_2 \left(\frac{K_2 - N_2 - a_{21} N_1}{K_2} \right) \quad (3)$$

Where K_1 , K_2 are the living environment loads of population N_1 , N_2 without competition to each other, and r_1 , r_2 are individual's maximum increasing rates. a_{12} , a_{21} are competition coefficients, a_{ij} represents the suppression effect of individuals of population N_i from individuals of population N_j . In community comprising n different populations, competition equation can be formulated as follows (4):

$$\frac{dN_i}{dt} = r_i N_i \left(\frac{K_i - N_i - \left(\sum_{j=1}^n a_{ij} N_j \right)}{K_i} \right) \quad (4)$$

It's the cooperation model based on ecological population density. We could exploit this model to describe the relation of multiple objectives. Because the relation of multiple objectives is just collaborative coexistence, and it finally is stable.

Exploiting (4), we proposed a ECCA algorithm based on ecological cooperation. The primary design goal of the proposed approach is to produce a reasonably good approximation of the true Pareto front of a problem. The proposed algorithm using dynamical equation of population competition at ecology to describe the complex, nonlinear relations of multiple objectives and to adjust the relation on individual and population levels simultaneously.

In our algorithms, each objective corresponds to a population. At each generation, these populations compete among themselves. An ecological population density competition equation is used for reference to describe the relation between multiple objectives and to direct the adjustment over the relation at individual and population levels. Moreover, the proposed approach store the Pareto optimal point obtained along the evolutionary process into external set, enforcing a more uniform distribution of such vectors along the Pareto front.

The basic idea of algorithm is as follows:

1. Each objective corresponds to a population;
2. In one iterative step, evolution process and cooperation process must be executed; the evolutionary process adopts GA's genetic operations, while the cooperation process adopts (4) to compute population density and to adjust the scales of populations. The scale of population is formulated as (5):

$$N_i(t+1)=N_i(t)+dN_i/dt \quad (5)$$

If the increasing of population N_i is positive, randomly generated dN_i/dt individuals join population N_i for enlarging the scale of N_i .

If the increasing of population N_i is negative, according to the fitness of population N_i , dN_i/dt individuals with minimal fitness are deleted. The scale of population is reduced.

As a complete unit, ECCA pseudo code description is given here:

Step1: Initialize a null set as external set

Step2: for all objective functions $f_i(x)$

Initialize a random population N_i corresponding to $f_i(x)$

endfor

Step3: while (terminative condition is NOT satisfied)

for all populations N_i

(A) Compute all of the Pareto optimal points P_i in population N_i

(B) Store P_i into external set

(C) General genetic operations are performed

(D) Computing dN/dt using (4)

(E) Determine next generation's scale of population N_i using (5)

endfor

endwhile.

3 Results of Simulation

To validate our approach, we used the methodology normally adopted in the evolutionary multiobjective optimization literature [8]. We performed quantitative comparisons (adopting three metrics) with respect to three MOEAs that are representative of the state-of-the-art in the area: the microGA for multiobjective optimization [9], the Pareto Archived Evolution Strategy (PAES) [10] and the Nondominated Sorting Genetic Algorithm II (NSGA-II) [11]. For our comparative study, we implemented for three following metrics:

1. Spacing (SP): This metric was proposed by Schott [12] as a way of measuring the range (distance) variance of neighboring vectors in the Pareto front known. This metric is defined as (6):

$$SP = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\bar{d} - d_i)^2} \quad (6)$$

Where $d_i = \min_j (\sum_{k=1}^m |f_m^i - f_m^j|)$, $i, j = 1, 2, \dots, n$, m is the number of objectives, \bar{d} is the mean of all d_i , and n is the number of vectors in the Pareto front found by the algorithm being evaluated. A value of zero for this metric indicates all the nondominated solutions found are equidistantly spaced.

2. Generational Distance (GD): The concept of generational distance was introduced by Van Veldhuizen & Lamont [13] as a way of estimating how far are the elements in the Pareto front produced by our algorithm from those in the true Pareto front of the problem. This metric is defined as (7):

$$GD = \frac{\sqrt{\sum_{i=1}^n d_i^2}}{n} \quad (7)$$

where n is the number of nondominated vectors found by the algorithm being analyzed and d_i is the Euclidean distance (measured in objective space) between each of these and the nearest member of the true Pareto front. It should be clear that a value of $GD=0$ indicates that all the elements generated are in the true Pareto front of the problem. Therefore, any other value will indicate how "far" we are from the global Pareto front of our problem.

3. Error Ratio (ER): This metric was proposed by Van Veldhuizen [14] to indicate the percentage of solutions (from the nondominated vectors found so far) that are not members of the true Pareto optimal set as (8):

$$ER = \frac{\sum_{i=1}^n e_i}{n} \quad (8)$$

where n is the number of vectors in the current set of nondominated vectors available; $e_i = 0$ if vector i is a member of the Pareto optimal set, and $e_i = 1$ otherwise. It should then be clear that $ER=0$ indicates an ideal behavior, since it would mean that all the vectors generated by our MEA belong to the Pareto optimal set of the problem. This metric addresses the third issue from the list previously provided.

We test the performance of ECCA on test function defined as follows (9):

$$\begin{aligned}
 & \text{Minimize } f_1(x_1, x_2) = x_1 \\
 & \text{Minimize } f_2(x_1, x_2) = (1.0 + 10.0x) \cdot \\
 & \left(1.0 - \frac{x_1}{1.0 + 10.0x} - \frac{x_1}{1.0 + 10.0x} \sin(\pi 4 x_1) \right) \\
 & 0.1 \leq x_1, x_2 \leq 1.0
 \end{aligned} \tag{9}$$

To compute the nondominated front for the ECCA, we did the following. For each ECCA run, we collected all the Pareto optimal point into external nondominated set corresponding to the individuals evaluated during the run.

In this example, our approach used: $\text{popsize}_{\text{init}} = 100$, $\text{popsize}_{\text{external}} = 30$. Table 1 shows the values of the metrics for each of the MOEAs compared.

As noted in the literature [15], comparing multiobjective optimization algorithms against each other can be difficult. One would like an algorithm to minimize the distance to the Pareto optimal front and provide uniform coverage of the Pareto optimal front for a wide range of values. Thus, comparisons become multiobjective optimization problems themselves: is an algorithm that finds a handful of Pareto optimal solutions better than an algorithm that finds a wide, uniform distribution of near Pareto optimal solutions? With this in mind we present the experimental results according to different algorithms shown in Table 1.

Table 1. Results of Simulation

		CO-MOEA	MicroGA	PAES	NSGAI
ER	best	0.46	0.42	0.02	0.00
	median	0.61	0.77	0.07	0.02
	worst	0.68	0.98	0.15	0.08
	average	0.60	0.75	0.07	0.03
	std. dev.	0.061	0.145	0.030	0.021
GD	best	0.0003	0.0008	0.0001	0.0007
	median	0.001	0.0089	0.0006	0.0008
	worst	0.042	0.238	0.0659	0.0009
	average	0.0049	0.0681	0.0066	0.0008
	std.dev.	0.009	0.086	0.016	0.000
SP	best	0.006	0.017	0.007	0.006
	median	0.012	0.042	0.014	0.008
	worst	0.379	1.539	0.624	0.086
	average	0.039	0.356	0.054	0.01
	std.dev.	0.073	0.507	0.141	0.014

From Table 1, we can see that the NSGA-II had the best overall performance. It is also clear that the microGA presented the worst performance for this test function. Based on the values of the ER and SC metrics, we can conclude that our approach had problems to reach the true Pareto front of this problem. Note however, that the values of GD and SP indicate that our approach converged very closely to the true Pareto front and that it achieved a good distribution of solutions. PAES had a good performance regarding closeness to the true Pareto front, but its performance was not so good regarding uniform distribution of solutions.

4 Conclusion

The solution of multiobjective optimization problem is a set of indeterminacy points, and the task of multiobjective optimization is to estimate the distribution of this solution set, then to find the satisfying solution in it [16]. Many methods solving multiobjective optimization using genetic algorithm have been proposed in recent twenty years. But these approaches tend to work negatively, causing that the population converges to small number of solutions due to the random genetic drift [17]. To avoid this phenomenon, a competitive coevolutionary genetic algorithm (ECCA) for multiobjective optimization is proposed. The primary design goal of the proposed approach is to produce a reasonably good approximation of the true Pareto front of a problem. In the algorithms, each objective corresponds to a population. At each generation, these populations compete among themselves. An ecological population density competition equation is used for reference to describe the relation between multiple objectives and to direct the adjustment over the relation at individual and population levels.

The proposed approach was validated using typical test function taken from the specialized literature. Our comparative study showed that the proposed approach is competitive with respect three other algorithms that are representative of the state-of-the-art in evolutionary multiobjective optimization. More work, and more comparisons is need to determine the general properties of ECCA, and how they can be adapted or improved.

References

1. Zitzler, E.: Multiobjective evolutionary algorithms: A comparative case study and the strength pareto approach. *IEEE Trans. on Evolutionary Computation* 13(4), 257–271 (1999)
2. Fonseca, C.M., Fleming, P.J.: An overview of evolutionary algorithms in multiobjective optimization. *Evolutionary Computation* 23(4), 1–16 (1995)
3. Zitzler, E., Deb, K., Thiele, L.: Comparison of Multiobjective Evolutionary Algorithms: Empirical Results. *Evolutionary Computation* 8(2), 173–195 (2000)
4. Ficici, S.G.: Solution concepts in coevolutionary algorithms, Ph.d. dissertation, Brandeis University (May 2004)
5. Hillis, W.D.: Co-evolving parasites improve simulated evolution as an optimization procedure. In: *Artificial Life II*, pp. 313–324. Addison-Wesley, Reading (1991)
6. Paredis, J.: Coevolutionary computation. *Artificial Life* 21(4), 355–379 (1995)

7. Cao, X.-b., Li, J.-l., Wang, X.-f.: Research on Multiobjective Optimization based on Ecological cooperation. *Journal of Software* 12(4), 521–528 (2001)
8. Coello Coello, C.A., Van Veldhuizen, D.A., Lamont, G.B.: *Evolutionary Algorithms for Solving Multi-Objective Problems*. Kluwer Academic Publishers, New York (2002); ISBN 0-3064-6762-3
9. Coello Coello, C.A., Toscano Pulido, G.: Multiobjective Optimization using a Micro-Genetic Algorithm. In: Spector, L., et al. (eds.) *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2001)*, pp. 274–282. Morgan Kaufmann Publishers, San Francisco (2001)
10. Knowles, J.D., Corne, D.W.: Approximating the Nondominated Front Using the Pareto Archived Evolution Strategy. *Evolutionary Computation* 8(2), 149–172 (2000)
11. 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)
12. Van Veldhuizen, D.A., Lamont, G.B.: On Measuring Multiobjective Evolutionary Algorithm Performance. In: *2000 Congress on Evolutionary Computation*, vol. 1, pp. 204–211. IEEE Service Center, Piscataway (2000)
13. Van Veldhuizen, D.A., Lamont, G.B.: On Measuring Multiobjective Evolutionary Algorithm Performance. In: *2000 Congress on Evolutionary Computation*, vol. 1, pp. 204–211. IEEE Service Center, Piscataway (2000)
14. Zitzler, E., Deb, K., Thiele, L.: Comparison of Multiobjective Evolutionary Algorithms: Empirical Results. *Evolutionary Computation* 8(2), 173–195 (2000)
15. De Jong, E.D., Pollack, J.B.: Idea evaluation from coevolution. *Evolutionary Computation* 12(4), 25–28 (2004)
16. Xing, Z.-Y., Zhang, Y., Hou, Y.-L., Jia, L.-M.: On Generating Fuzzy Systems based on Pareto Multi-objective Cooperative Coevolutionary Algorithm. *International Journal of Control, Automation, and Systems* 5(4), 444–455 (2007)
17. Horn, J., Goldberg, D.: A Niche Pareto genetic algorithm for multiobjective optimization. In: *Proc. of the First IEEE Conference on Evolutionary Computation*, Piscataway, pp. 82–87 (1994)

Secure Secret-Key Management of Kerberos Service

Lai-Cheng Cao*

School of Computer and Communication
Lanzhou University of Technology
Lanzhou 730050, China
caolch@lut.cn

Abstract. The Kerberos protocol has promoted the development of new techniques to support various kinds of distributed applications. However, the secret-key management is security core in the whole system. Using symmetric encryption algorithm Rijndael of AES (Advanced Encryption Standard), all secret-keys of the client were encrypted by the secret-key of the authentication server and stored in the database. The secret-key of the authentication server was protected by distributing its shares to the router, Ticket-granting Server (TGS) and the Web server. The authentication server did not store its secret-key in system, when the system needed this secret-key, the authentication server could synthesize it by distributed shares. Security analysis shows that this secret-key management has fault-tolerant and no-information leakage; it also defends collusive attack and cracking the secret-key attack.

Keywords: The secret-key, Kerberos protocol, Rijndael encryption algorithm, fault-tolerant.

1 Introduction

The Kerberos protocol developed by MIT has been widely used in the security authentication service for users and network connection. It bases on symmetric encryption algorithm, the communication uses share-secret-key, and these share-secret-keys of clients and server are stored into the database. Once the intruder unauthorized accesses the database, all secret-keys can be leaked, there is no security to speak of the whole system. The paper [1] extends an ideal functionality for symmetric and public-key encryption proposed in previous work by a mechanism for key derivation. Ke Jia etc [2] propose the public key encryption algorithm based on braid groups to improve Kerberos protocol. Liu ke-long etc [3] present an improved way using Yaksha security system of ElGamal Public Key algorithm based on the original protocol framework to overcome the Kerberos' limitations. All these methods have extended public key algorithm based on symmetric encryption algorithm, it makes key management more complex and system speed slower. In this paper, we put

* This work is supported by the National Natural Science Foundation of China under Grant No. 60972078; the Gansu Provincial Natural Science Foundation of China under Grant No. 0916RJZA015.

forward scheme to protect the secret-keys of the client's users based on symmetric encryption algorithm Rijndael of AES, all secret-keys except the identity (ID) of users are encrypted by the secret-key of the authentication server and stored in the database. The secret-key of the authentication server is protected by distributing its shares to the router (or the gateway), Ticket-granting Server (TGS) and the Web server.

The remainder of this work is organized as follows. In section 2 we describe the Kerberos protocol. In Section 3 we present a scheme to protect the secret-key of the users. In section 4 point out a scheme to protect the secret-key of the authentication server. In section 5 we present the correctness proof of this scheme. In section 6 we finish security analysis about our method. The conclusion is presented in section 7.

2 The Kerberos Protocol

In the distributed environment, in order to protect user information and server resources, we require that client systems authenticate themselves to the server, but trust the client system concerning the identity of its user. We describe the Kerberos protocol [4-6] (as shown in Fig. 1); it includes the gateway, Authentication Server (AS), Ticket-granting Server (TGS), web server. Authentication Server keeps a database containing the secret-keys of the users and the router and TGS. The client and the Server can communicate each other based on TCP/IP protocol.

- *Client*: Requires gaining access to Web server in the distributed environment.
- *AS*: Authenticates servers to the client.
- *TGS*: Grants service-granting ticket to the client.
- *Web server*: Stores resource and data for the web users.

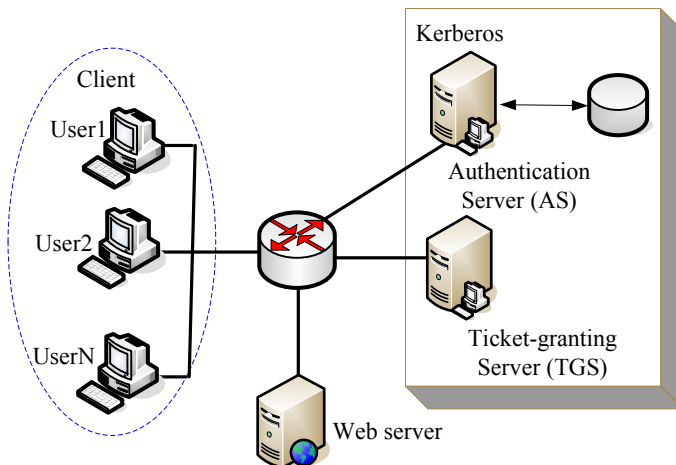


Fig. 1. The architecture about the Kerberos protocol

If the client requests to access Web server, this server will authenticate the client. Similarly, if Web server requests to access client, this client authenticates web server.

3 Protecting the Secret-Key of the Users

All secret-keys of the users, the secret-keys of TGS and the router are encrypted by the secret-key of the authentication server based on AES encryption algorithm. Unencrypted users' ID and their encrypted secret-key, unencrypted TGSs' ID and its encrypted secret-key, unencrypted router's ID and its encrypted secret-key are stored in the database; this database table is showed in Table 1 as follows.

Table 1. Encrypted secret-key OF database table

Primary key	Encrypted secret-key	Index
<i>TGS_id</i>	*****	<i>1</i>
<i>Router_id</i>	*****	<i>2</i>
<i>User1_id</i>	*****	<i>3</i>
<i>User2_id</i>	*****	<i>4</i>
.....
<i>UserN_id</i>	*****	<i>N+2</i>

When the system needs a secret-key, first it searches the ID of this secret-key in this database table, and then decrypt this encrypted secret-key with the secret-key of the authentication server. Of course, the security of the secret-key of the authentication server is a key and core of security of entire system, thus, we put forward that the shares of the secret-key of the authentication server are distributed to store in the router (or the gateway), Ticket-granting Server (TGS) and the Web server, and the authentication server does not save its secret-key, when the authentication server needs this secret-key, it can be synthesized by these shares, and the authentication server throws away it after it is used.

4 Protecting Secret-Key of the Authentication Server

4.1 Generating Pre-datum of the Shares

Referring to the paper [7-8], we divide the secret-key of the authentication server into three shares, they are stored in four share servers, these four share servers are the authentication server, the router, Ticket-granting Server (TGS) and the Web server. The secret-key of the authentication server can be synthesized by three shares of three servers which are selected from four servers, the shares are designed for two groups to increase fault tolerance. The authentication server produces pre-datum of the share before these shares are generated. Table 2 shows these pre-datum which are belong to four share servers, d_{11} and d_{21} are pre-datum which shall generate the shares of the authentication server, and they each belong to group 1 and group 2, d_{12} and d_{21} are pre-datum which shall generate the shares of the router, and they each belong to group

1 and group 2, d_{13} and d_{22} are pre-datum which shall generate the shares of TGS, and they each belong to group 1 and group 2, d_{13} and d_{23} are pre-datum which shall generate the shares of the Web server, and they each belong to group 1 and group 2. These pre-datum can compound: $d=d_{11}+d_{12}+d_{13}=d_{21}+d_{22}+d_{23}$.

Table 2. The pre-datum of the share belong to share server

Group number	Authentication server	Router	TGS	Web server
Group 1	d_{11}	d_{12}	d_{13}	d_{13}
Group 2	d_{21}	d_{21}	d_{22}	d_{23}
Compounding 1: $d=d_{11}+d_{12}+d_{13}$				
Compounding 2: $d=d_{21}+d_{22}+d_{23}$				

4.2 The Method of Distributing Shares

Referring to the paper [9-10], the method of distributing shares is presented the follow.

1) Choosing a polynomial:

$$f(x) = \sum_{k=0}^3 a_k x^k \quad (1)$$

2) Generating 4 random numbers x_i ($i=0,1,\dots,3$), and computing:

$$f(x_i) = \sum_{k=0}^3 a_k x_i^k$$

3) When $i=0$, taking $d_0 = f(x_0)$ as the pre-datum of synthesizing share of the authentication server.

4) When $i=1,2,3$, using LaGrange formula to computer:

$$f(x) = \sum_{i=1}^3 (f(x_i) \sum_{j=1,3}^{j \neq i} \frac{x - x_j}{x_i - x_j}) \quad (2)$$

$$\text{Let } x=0: f(0) = \sum_{i=1}^3 (f(x_i) \sum_{j=1,3}^{j \neq i} \frac{-x_j}{x_i - x_j}) \quad (3)$$

5) Taking m ($m=1,2$) group pre-datum of the shares:

$$d_{mi} = f(x_i) \sum_{j=1,3}^{j \neq i} \frac{-x_j}{x_i - x_j} \quad (i=1,2,3) \quad (4)$$

6) Computing pre-datum d of the shares:

$$d = d_0 + \sum_{i=1}^3 d_{mi} \quad (5)$$

7) Then the authentication server distributes pre-datum d_{mi} ($m=1,2; i=1,2,3$) of the shares to the share servers.

4.3 Generating the Shares of Share Servers

We adopt the computational infeasibility of discrete logarithms to generate the shares of share servers. Table 3 shows the group of these shares S_{mi} ($m=1, 2; i=1, 2, 3$). When any one share server loses its shares, we can adopt the remaining three share servers to synthesize the secret-key, and the system can continuously work. When any two share servers are attacked, any information of the secret-key can not be leaked.

Table 3. The key share of 3 server-from-4 server

Group number	Authentication server	Router	TGS	Web server
Group 1	S_{11}	S_{12}	S_{13}	S_{13}
Group 2	S_{21}	S_{21}	S_{22}	S_{23}

The whole system chooses two big prime number p and q .

1) The authentication server selects an integer $x_k \in [1, p-1]$ and computes:

$$y_k = p^{x_k} \text{ mod } q \tag{6}$$

2) The authentication server broadcasts y_k to the share servers.

3) The share servers compute own shares:

$$S_{mi} = y_k^{d_{mi}} \text{ mod } q \quad (m=1, 2; i=1, 2, 3) \tag{7}$$

4.4 Synthesizing the Secret-Key of the Authentication Server

When the authentication server needs its secret-key, it selects one group shares whose numbers are m_1, m_2, m_3 ($m=1, 2$) from three of the four servers, and then synthesize the secret-key as follows.

1) The authentication server computes:

$$S_0 = y_k^{d_0} \text{ mod } q \tag{8}$$

2) When the authentication server receives the shares of share server, it

synthesizes its secret-key: $\text{Secret-Key} = S_0 \prod_{i=1}^3 S_{mi}$ (9)

The left figure in Fig. 2 shows that the authentication server uses the first group shares to synthesize the secret-key based on the authentication server, the router and TGS. The right figure in Fig. 2 shows same result based on the authentication server, the router and the Web server.

Where, $\text{Secret - Key} = S_0 \prod_{i=1}^3 S_{1i}$

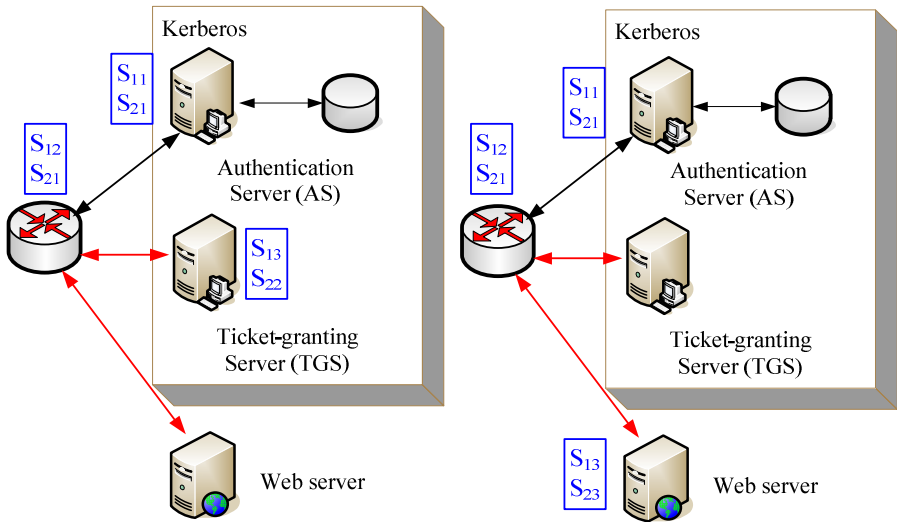


Fig. 2. Synthesizing the secret-key by first group shares

The left figure in Fig. 3 shows that the authentication server uses the second group shares to synthesize the secret-key based on the router, TGS and the Web server. The right figure in Fig. 3 shows same result based on the authentication server, TGS and the Web server. Where, Secret-Key = $S_0 \prod_{i=1}^3 S_{2i}$

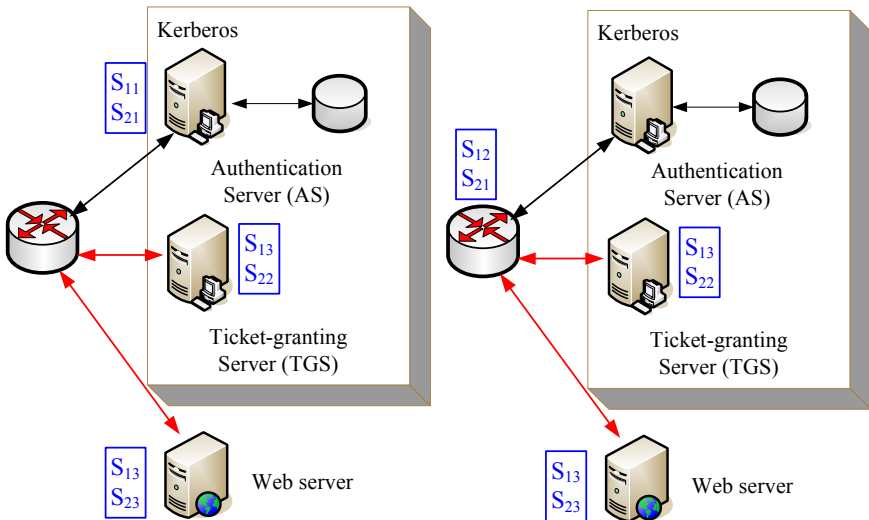


Fig. 3. Synthesizing the secret-key by the second group shares

5 Correctness Proof

Let d_{mi} and d_{li} ($m, l=1, 2; i=1, 2, 3$) as two group pre-datum, their shares are S_{mi} and S_{li} . K_m and K_l are synthesized secret-key based on two group shares.

Because $d = d_0 + \sum_{i=1}^3 d_{mi} = d_0 + \sum_{i=1}^3 d_{li}$, thus

$$K_m = S_0 \prod_{i=1}^3 S_{mi} = y_k^{d_0} \text{ mod } q \prod_{i=1}^3 y_k^{d_{mi}} \text{ mod } q = y_k^{d_0 + \sum_{i=1}^3 d_{mi}} \text{ mod } q = y_k^d \text{ mod } q$$

$$K_l = S_0 \prod_{i=1}^3 S_{li} = y_k^{d_0} \text{ mod } q \prod_{i=1}^3 y_k^{d_{li}} \text{ mod } q = y_k^{d_0 + \sum_{i=1}^3 d_{li}} \text{ mod } q = y_k^d \text{ mod } q$$

Therefore, $K_m = K_l$.

6 Security Analysis

The security of this secret-key management of Kerberos service can be attributed to four factors:

1) Fault-tolerant

According to equation (7) and (9), when one share sever loses its shares, the system can continuously work.

2) No-information leakage

Because equation (6), (7) and (8) base on the computational infeasibility of discrete logarithms, only $y_k = p^{x_k} \text{ mod } q$ and $S_{mi} = y_k^{d_{mi}} \text{ mod } q$ can reflect the information of the secret-key, it is impossible that using y_k and S_{mi} to compute d_{mi} and x_k , namely, attacker can not obtain any information of d_{mi} and x_k from filched S_{mi} and y_k .

3) Defending collusive attack

If three ones of four share severs perform collusive attack, they can select one group shares whose share numbers are sequential after they are arrayed again, but d_0 and d_{mi} ($i=1,2,3$) are random and irrelative, and conditional information entropy $H(d_0 | d_{mi}) = H(d_0)$, in addition, participating x_k in synthesizing the secret-key is randomly selected in [1, p-1], therefore, collusive attack can not finish.

4) Resisting cracking the secret-key attack

Rijndael encryption algorithm uses secure S-boxes as nonlinear components [9, 10]; it supports on-the-fly subkey computation for encrypting. The operations used by Rijndael are among the easiest to defend against power and timing attacks. The use of masking techniques to provide Rijndael with some defense against these attacks does not cause significant performance degradation. So it highly resists cracking the secret-key attack.

7 Conclusion

The secret-key management is security core in the Kerberos protocol; it is also a hot research field of information security. In this paper, we adopt encryption algorithm Rijndael of AES, distributing shares of the secret-key and synthesizing the secret-key, to protect effectively the secret-key. This technology makes the whole system to have fault-tolerant and no-information leakage, and also defends collusive attack and cracking the secret-key attack.

References

1. Küsters, R., Tuengerthal, M.: Ideal Key Derivation and Encryption in Simulation-Based Security. In: Kiayias, A. (ed.) CT-RSA 2011. LNCS, vol. 6558, pp. 161–179. Springer, Heidelberg (2011)
2. Jia, K., Chen, X., Xu, G.: The improved public key encryption algorithm of Kerberos protocol based on braid groups. In: 2008 International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM 2008), vol. 1, pp. 1–4 (2008)
3. Liu, K.-l., Qing, S.-h., Yang, M.: An Improved Way on Kerberos Protocol Based on Public-Key Algorithms. *Journal of Software* 12(6), 872–877 (2001)
4. Lai-Cheng, C.: Enhancing distributed web security based on kerberos authentication service. In: Wang, F.L., Gong, Z., Luo, X., Lei, J. (eds.) *Web Information Systems and Mining*. LNCS, vol. 6318, pp. 171–178. Springer, Heidelberg (2010)
5. Rao, G.S.V.R.K.: Threats and security of Web services - a theoretical short study. In: *Proceedings of IEEE International Symposium Communications and Information Technology*, vol. 2(2), pp. 783–786 (2004)
6. Seixas, N., Fonseca, J., Vieira, M.: Looking at Web Security Vulnerabilities from the Programming Language Perspective: A Field Study. *Software Reliability Engineering* 1, 129–135 (2009)
7. Wu, T., Malkin, M., Boneh, D.: Building intrusion-tolerant applications. In: *Information Survivability Conference and Exposition*, pp. 25–27. IEEE Computer Society, Los Alamitos (2000)
8. Zhang, X.-f., Liu, J.-d.: A threshold ECC Based on Intrusion Tolerance TTP Scheme. *Computer Applications* 24(2), 5–8 (2004)
9. Zhendong, S., Gary, W.: The essence of command injection attacks in web applications. *ACM SIGPLAN Notices* 41(1), 372–382 (2006)
10. Ashley, C., Wanlei, Z., Yang, X.: Protecting web services from DDOS attacks by SOTA. In: *ICITA 2008*, pp. 379–384 (2008)

Application of Shuangchao Hydrological Model in Shanxi Semiarid and Semi-humid Area

Yutao Cao and Xihuan Sun

College of Water Resources Science and Engineering
Taiyuan University of Technology, TYUT
Taiyuan, China
caoyutao@tyut.edu.cn

Abstract. At present, the watershed hydrology model in semiarid and semi-humid areas do less work whether in foreign or in our country, therefore studying on the hydrology simulation in semiarid and semi-humid area is very important. This paper focuses on the principle and the structure of Shuangchao hydrological model, which was applied to simulate 5 historical floods of Zhangfeng Reservoir in Shanxi, and the results show the model is very suitable for this basin. Finally, the author puts forward some suggestions for future improvement of Shuangchao hydrological model in the future.

Keywords: Flood forecasting, Shuangchao Hydrological Model, semiarid and semi-humid.

1 Introduction

Hydrological phenomenon of the nature is a complex process, which is interacted with multiple factors, and because of its formation mechanism is not entirely clear, hydrological model becomes into a very important tool for studying the complex hydrological phenomenon and hydrological forecasting. The generation of hydrological model is the inevitable result in studying the law of the hydrological cycle. Hydrological model belongs to mathematical model, it uses mathematical methods to describe and simulate the process of the hydrological cycles, explains some experience laws by physics, expresses by a series of rigorous mathematical equations, and at last integrates all hydrological together to form the water balance calculation system of the basin[1-2].

The forecasting is an important part of the hydrological work, in the past few decades, it has great development and has accumulates wealth of experience. Especially recently, with computers, networks, remote sensing, geographic information technology applying in the hydrological forecasting, the progress of hydrological forecasting has new development. However, in the development work of in hydrological model, there are so much works doing in the south humid areas, and calculation methods are more mature, while the work in semi-dry areas has done less. Whether foreign or domestic hydrological models, when they are used in semiarid and semi-humid areas, there exist problems such as unknowing the flood courses, low

prediction accuracy, which can not be good for the drought, water diversion, mountain hazards control, water environment manage and hydrology projects optimize to provide technical services.

Semi-dry areas of our country are mainly distributed in north, northeast and central and eastern Tibetan Plateau, and the semi-dry areas accounting for about 52% land area in China, therefore, it has very important in studying of the semi-dry areas' hydrological situation and it is very significant to establish the region hydrological model.

In this paper, the author mainly studies the application of Shuangchao hydrological model at Zhangfeng Reservoir in Shanxi. Through the analysis of the model, the author is going to further explore the rainfall runoff characteristics in the semiarid and semi-humid area, and proposed recommendations of the model for future development.

2 Shuangchao Hydrological Model

In the semiarid and semi-humid area, because of less precipitation, the soil often in the status of lacking water, and its runoff producing and confluence process characteristics are different with the humid areas. The high intensity short time rainfall will excess to generate surface runoff after long time drought, and the volume of the water infiltration is not so much at this time. The infiltration front is often difficult to reach the less permeable layer, upper water will not exceed and has no lateral flow of water, and thus it can not produce interflow. When it happened to the rainfall or the behind of the long period precipitation after long time rain, the interflow will appear since the above aquitard interface layer exceed holding water. According to the semiarid and semi-humid area runoff mechanism, Senior Engineer Wang Yinjie, whom works in Hydrology and Water Resources Survey Bureau of Shanxi Province, bring forward the Shuangchao hydrological model. In this model, excess infiltration will produce surface runoff; excess holding will produce interflow and underground runoff. The Shuangchao model is drawn on the design ideas by the Shanbei hydrological model, Xin'anjiang hydrological model, Janpan Tank model, the United States Sacramento model and other models [3-4].

Shuangchao runoff model thinks that: as soon as the rainfall intensity exceeds infiltration capacity, then it becomes the surface runoff, which is an integral part of the runoff, but not all. The precipitation, which penetrates into the unsaturated zone or the influence layer, will move along with the interface of the layered media when it reaches to water holding capacity and forms the gravity water. And when encountering with gully or foot of the river slope or slopes of the saturated zone, it will escape and form interflow, which its proportion in total runoff is determined by the characteristics of the precipitation and land surface in the basin. The structure of the model is designed according to the physical mechanism of three kinds of water and includes five components: fictitious infinitesimal infiltration, watershed distribution of the infiltration capacity, surface runoff, interflow and groundwater runoff, soil evaporation and soil moisture.

The basin confluence runoff model is taken advantage of hysteresis Nash model, and is used water storage balance and discharge equation to solve the differential equation of Nash model, $s(t)$ curves, and period confluence curve. River confluence model is adopted by diffusion simulations method, which is the hydraulic linear analytical solution of diffusion wave for the Saint-Venant equations. It can be used for the channel which the bottom width is B , the average water depth is D , the wide-shallow water surface down is SW and the channel is wide-shallow prism and without water seepage fill.

Shuangchao hydrological model structure is shown in Figure 1 [5-11].

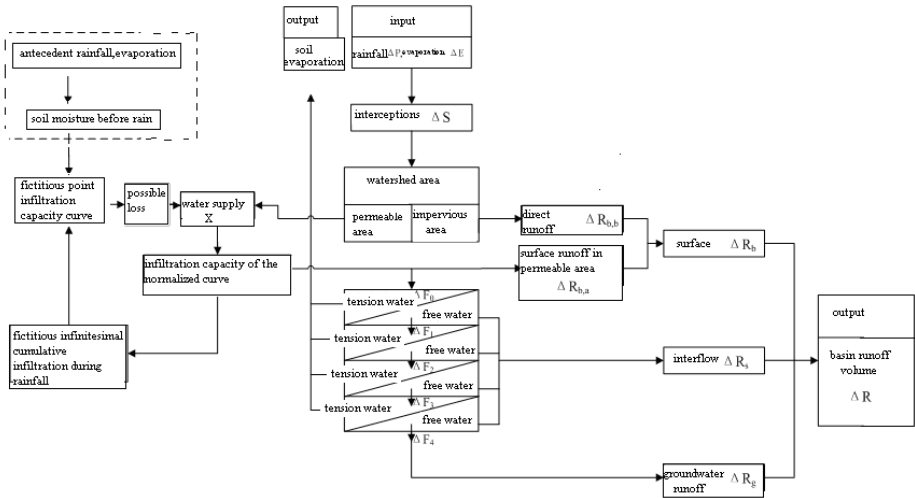


Fig. 1. Shuangchao hydrological model flow chart

3 Engineering Background and Application

3.1 Engineering Background

Zhangfeng Reservoir is located in Village Qinhe River in Zhengzhuang town Qinshui County of Jincheng City, Shanxi Province, and the distance to the Jincheng City is 90km. The reservoir control drainage area is 4990km², the total capacity of the reservoir is 394 million m³, it is a large scale hydro project which undertaking the main object of water supply to urban living and industrial, and the secondary object is flood control, power generation and other utilization task. The hydro project is built up by the dam, diversion tunnel, spillway, water supply power tunnel and headwork pumping station and other components. The average annual precipitation in the basin is between 550 ~ 583.9mm, and the precipitation proportion in flood season (June to September) is 67 to 74%of the total annual. The average annual evaporation is 1517 to 1827.8mm, and the designed years average runoff is 479 million m³, annual average flow is 15.2m³/s.

Zhangfeng reservoir engineering started to build on November 17, and the main project, that is the weir dam, began to build in June 2005. The dam is clay slope core rock fill dam, the maximum height is 72.2m, normal water level is 759m. Cofferdam High 23.24m, crest elevation of 725.24m, the design flood standard of once in 10 years, and the corresponding high flood water level is 724.04m.

Hydrological flood forecasting station network of Zhangfeng reservoir includes: 4 hydrological station such as Kongjiapo, Feiling, Youfang, Zhangfeng and 19 telemetry rainfall station between Feiling and Zhangfeng. Among the 23 telemetry station, Dajiang, Weizhai and Longqu observe both the water level and flow.

3.2 Application of Shuangchao Model in Zhangfeng Reservoir

The flood forecast program of Zhangfeng reservoir is as follows: according to the river inflow course of the Feiling hydrological station and the interval precipitation between Feiling and Zhangfeng, and river confluence runoff model, basin rainfall runoff and confluence model to calculate the flood process of the Zhangfeng reservoir. The method utilizes sub-unit, sub-perion and sub-water to calculate runoff and confluence process of every unit of the basin. The runoff model uses Shuangchao model, the basin confluence model uses Nash model and the river confluence uses diffusion simulation to compute. In order to better control the every tributary's rainfall runoff, the every tributary is thought as a unit. Thus the basin of Zhangfeng is divided into 9 units and every unit calculates lonely. Basin unit is shown in Figure 2 and Table 1.

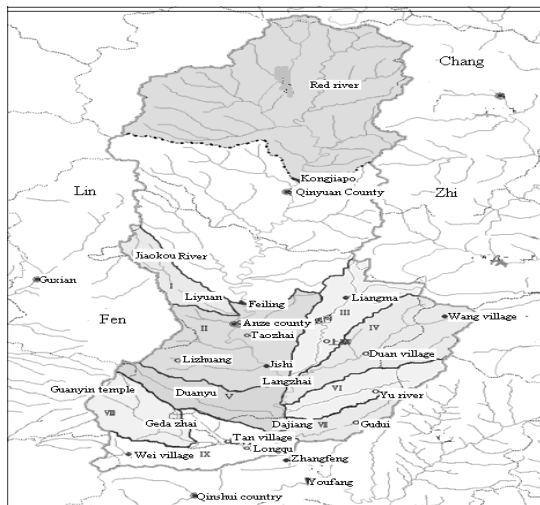


Fig. 2. Forecasting unit of Zhangfeng reservoir

Table 1. Unit area and rainfall station tables of Zhangfeng reservoir

Unit	Area (km ²)	Section of river confluence	Length from the section before (km)	Rainfall station
I	162	Weizhai	41.6	Jiaokouhe,Liyuan
II	491	Weizhai	0	Lizhuang,Feiling,Jishi,Taozhai
III	262	Weizhai	0	Shangzhai,Liangma
IV	358	Weizhai	0	Wang village,Du willage
V	219	Dajiang	0	Duanyu,Dajiang
VI	195	Dajiang	0	Weizhai,Wang river
VII	130	Zhangfeng	0	Dajiang,Gudui
VIII	177	Longqu	28.6	Zhaigeda,Guanyin temple
IX	313	Longqu	0	Wei and Tan village,Longqu

We choose 10 typical historical floods in 7 years of Zhangfeng reservoir, 1988, 1992, 1993, 1996, 2001, 2003, and 2005. And the former 5 historical flood are used to calibrate the parameters of the model, the last 5 flood data are used to test model accuracy. Runoff depth forecast error is 20% of the measured, and when it is less than 3mm, the take 3mm as the error. The parameters of Shuangchao hydrologic model are list below from table 2 to table 4. And the last five runoff flood forecasting pass rate is 80%^[12]. The assessment of the flood results is in table 5.

Table 2. Parameters of runoff for each unit

Unit	Hydraulic conductivity ks	Diffusivity kr	Infiltration curve index b	Lateral discharge share coefficient δ	Critical rainfall intensity α_0
I	10	65	5	0.04	0.05
II	10	65	5	0.04	0.05
III	6	50	5	0.04	0.05
IV	6	50	5	0.04	0.05
V	10	80	5	0.04	0.05
VI	6	50	5	0.04	0.05
VII	6	50	5	0.04	0.05
VIII	20	100	5	0.12	0.1
IX	15	100	5	0.12	0.1

Table 3. Parameters of basin confluence

Unit		I	II	III	IV	V	VI	VII	VIII	IX
Surface	Displacement delay time τ	3	3	3	3	3	3	3	3	3
	Planarization delay time kr	3	5	3	4	3	3	3	3	4
	n	1	1	1	1	1	1	1	1	1
interflow	Displacement delay time τ	4	4	4	4	4	4	4	4	4
	Planarization delay time kr	7	7	7	7	7	7	7	7	7
	n	2	2	2	2	2	2	2	2	2

Table 4. Parameters of river course confluence

River course	Outlet of unit I to Weizhai	Weizhai to Dajiang	Outlet of unit VIII to Longqu	Dajiang to Zhangfeng
Diffusion coefficient μ	200	200	200	200
Flood velocity u	1.5	1.5	1.2	1.5
River length L (m)	41.6	20.2	28.6	17.6

Table 5. Assessment results of Zhangfeng reservoir historical flood

No.	Flood	Runoff depth		Error	Allowable Error	Fitness Or not
		Measured	Calculation			
1	19930805	8.4	10.7	2.3	3	√
2	19960731	45.9	43.6	-2.3	9.2	√
3	20010728	5.7	4.0	-1.8	3	√
4	20030827	21.6	24.5	2.9	4.3	√
5	20050921	11.4	22.2	10.8	3	×

Fig 3 and Fig 4 show two large flood event course simulations of historical flood.

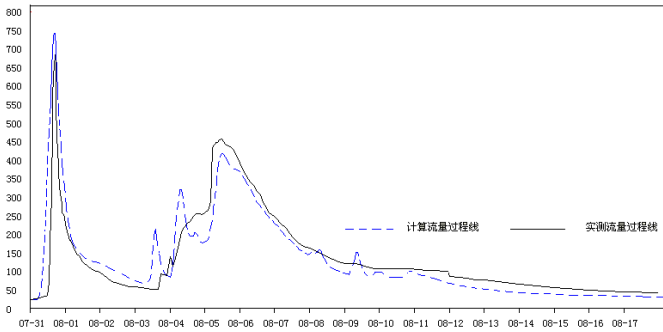


Fig. 3. 19960731

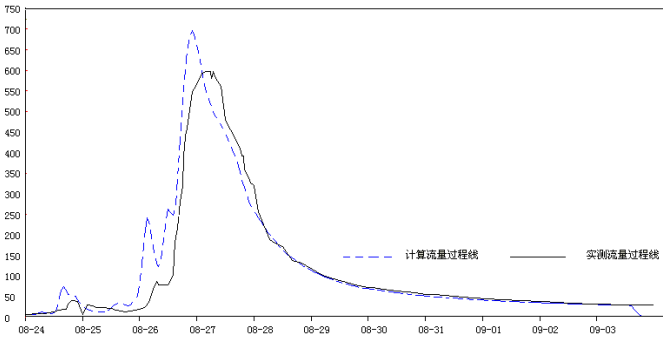


Fig. 4. 20030827

4 Conclusion

The core of Shuangchao hydrological model structure is to use runoff theory combined by infiltration curve and basin infiltration distribution curve, and simulate the exceed surface water. In this paper, five historical floods were predicted using Shuangchao hydrological model and the accuracy of forecasting results were very ideal, so we can see that the model has great applicability in Zhangfeng reservoir. However, as a relatively new hydrological model using in semiarid and semi-humid area, Shuangchao model is not perfect in some respects and it needs adjust in some place.

The model assumes that infiltration capacity distribution curve has the same distribution at any point in time; this generalization can not reflect changes in the basin excavated wetland distribution of infiltration capacity. Therefore this assumption is too simple and can not reflect the variation of the infiltration capacity; we can try to consider soil moisture changes in the model so that to improve the simulation accuracy. At the same time, considering to analysis with other hydrological models and use different models to simulate different reservoirs' flood in order to get the reasonable hydrological model in Shanxi semiarid and semi-humid area future.

Acknowledgment. The paper is supported by the Youth Fund of Taiyuan University of Technology 2010.

References

1. Wang, M., Zhang, Q.: The construction of E-business: a case study. *Journal of Peiking University* 243, 102–103 (2003)
2. Kachroo, R.K.: River Flow forecasting(Special issue). *Journal of Hydrology* 133, 1–5 (1992)
3. Renjun, Z.: *Hydrological Simulation*. Water Conservancy and Electric Power Press, Beijing (1984)
4. Nakagiri, T., Watanabe, T., Moruyama, T.: Performance of low flow prediction system. In: *Proc.of International Conference on Water Resources and Environment Research: Towards the 21st Century, Japan* (1996)
5. Wang, Y.: *Flood Forecasting Principles and Methods of training materials* (Unpublished draft) (2004)
6. Wang, Y.: Analysis of the unsaturated soil moisture function and new interpretation of the Richards infiltration equation. *Hydrology* 1996(2), 1–6 (1996)
7. Wang, Y.: Discussion on unsaturated soil of Richards's infiltration equation. *Hydrogeology and Engineering Geology* 2004(1), 9–13 (2004)
8. Hu, C.: Progress and Prospects of hydrological models. *South- to -North Water Transfers and Water Science & Technology* 2(6), 29–30 (2004)
9. Chen, Y.: Problems on flood forecasting in the semiarid region. *Advances in Water Science* 2003(5), 79–83 (2003)
10. Zhang, J.: A Study on Demacation Indexes between Subhumid and SemiaridSectors in China. *Porgress in Geography* 18(3), 230–237 (1999)
11. Ma, Z.: Decadal variations of arid and semi-arid boundary in China. *Chinese Journal of Geophysics* 48(3), 519–525 (2005)
12. Yiling, Z.: *Hydrological Forecast*. Water Conservancy and Electric Power Press, Beijing (1989)

The Design of Withstand Test Simulation System Based on Resonant Principle

Yonggang Li, Xia Ping, Chen Zhao, and Shuting Wan

School of Electrical and Electronic Engineering, North China Electric Power University,
071000 Baoding, China
495281727@qq.com

Abstract. Before the electric power equipments are in operation, the maintenance department of power system need to conduct a withstand test to them. The withstand test simulation system is based on the basic principle of withstand test. Through the operators' visual build to the power modules, the software can simulate digitally and calculate automatically for the test. So, researchers were able to manage the withstand test visually. It makes them to manage the work of withstand test more effectively.

Keywords: Electric power equipment, Withstand test, Visual build to the power modules, Digital simulation.

1 Introduction

Withstand test based on the resonant principle is the main mean to excute withstand tests for many electric power equipments, especially for cables. The operator need formulate a test plan according to the test equipment before the test. But in the actual work, the plan need a large amount of calculation and higher theoretical level. Because it can't transform the mode of connection flexibly, this will greatly influence the working efficiency.

In this system, the operator selects the variable-frequency power, the excitation transformer from the library of test equipment, and lap the resonator reactors manually to generate a complete resonant withstand-testing circuit. By digital simulation, operators can get the electrical information of the circuit and every electrical device very clearly before the test. So, they can manage the work of withstand test more effectively and easily.

2 Establishment of Electrical Model

The Modeling of withstand test circuit is the principal question of digital simulation. There are two fashions of series resonance including frequency modulation and reactance modulation. The fashion of reactance modulation uses the Reactor that can regulate core air gap. It regulates reactance to keep power frequency. The fashion of frequency modulation uses combination of fixed Reactor. It regulates the frequency of variable-frequency power sources to keep resonance. Then, the testing appears high

voltage to come true the goal of withstand voltage test. Comparing to the fashion of reactance modulation, the fashion of frequency modulation has so many advantages, such as minute extension, light weight, high quality factor ,easy and simple to handle, and so on, so it is widely applicable [1]. Therefore, this article uses the withstand voltage test circuit of frequency modulation as the electrical model. The circuit diagram of withstand test circuit of frequency modulation is shown in Figure 1.

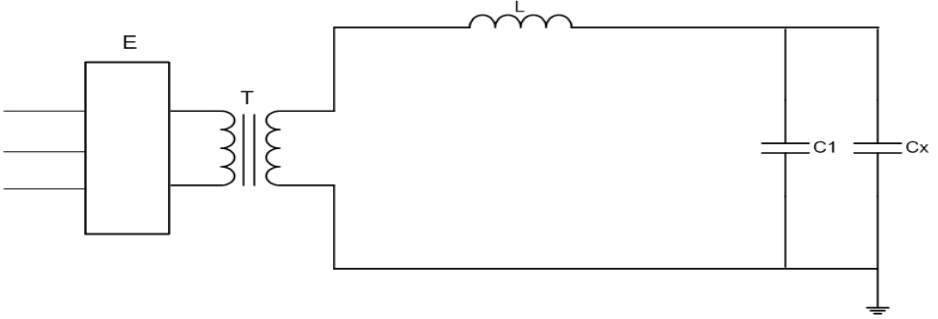


Fig. 1. The equivalent diagram of withstand test circuit

E—Variable-frequency power; T—Excitation transformer; L—Resonant reactor; C1—Compensation and voltage division capacitance; Cx—Equivalent capacitor of the product to be tested

The voltage between capacitor is:

$$U_c = I_s X_c = U_s X_c / \sqrt{R^2 + (X_L - X_c)^2} \quad (1)$$

Among them, I_s : Total current of the circuit, U_s : Output voltage of the excitation transformer, X_c : Total capacitance of the compensatory capacitor and the equivalent capacitor of the product to be tested, X_L : Total impedance of the resonator reactors, R : Total resistor of the circuit

If $X_L = X_c$, the series resonance occurs, and U_c achieve maximum, At this time :

$$1 / 2\pi fC = 2\pi fL \quad (2)$$

$$U_c = U_s X_c / R \quad (3)$$

The resonant frequency can be obtained as follows:

$$f = 1 / 2\pi\sqrt{LC} \quad (4)$$

The resonant current of circuit is:

$$I_s = U_s / \sqrt{R^2 + (X_L - X_c)^2} = U_s / R \quad (5)$$

When the resistance of products to be tested is ignored, the current of products to be tested is:

$$I_c = U_c / X_c = \omega C U_c = 2\pi f C U_c \quad (6)$$

When the withstand test does not consider the effects of corona on the circuit, the quality factor Q is:

$$Q = \sqrt{L / C} / R = \omega L / R = 1 / \omega C R \quad (7)$$

Then:

$$U_c = Q U_s \quad (8)$$

Namely ,the voltage of the product to be tested is the Q times excitation voltage. So we have achieved the purpose of conducting the withstand test to the large-capacity product with a small-capacity transformer [2].

3 The Design of the System

3.1 The Platform of the Development

The VB Language is concise, has a strong versatility, and don't need programmers to write much code to describe the appearance and location of interface element. With the simple drawing function and other comprehensive function, VB can satisfy the requirements of the research. So we choose Visual Basic 6.0 as the development tool. And we choose Microsoft Office Access as the backend database. Traditional database interfaces——Open Database Connection (ODBC) ——has problems of slow storage and complicated configuration. ActiveX Data Objects (ADO) is the good choice to connection backend database dynamically, to realize fast and efficient data exchanges and to improve the speed of the system [3].

3.2 General Design Thought

The system laps a complete circuit by inputting the test voltage and the capacitance of product, choosing the excitation transformer, the variable-frequency power and combining reactors manually. Among these, the core part is combining reactors manually. According to the parameters and combinations of the reactor, the system should compute the total inductance, resistance and the allowed-maximum current and voltage automatically. After running, the software should calculate the resonant frequency, the quality factor and the voltage and current of the components based on the parameters of the circuit. If the result exceed its allow ratings, the system will give alarms. When the total current is too high, it should have the function of adding parallel resonant branch to reduce the current [4]. The design flow chart shows in Fig 2.

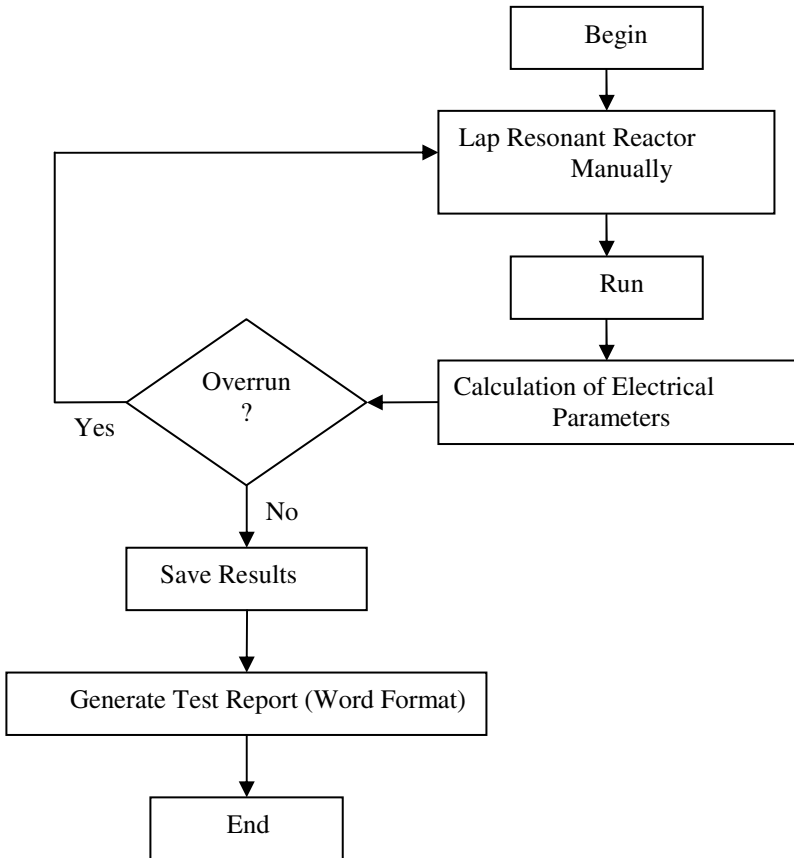


Fig. 2. The design flow chart

3.3 The Design of the Main Interface of the System

The sites of the variable-frequency power and the excitation transformer are fixed in the withstand test, so the system should lap them in advance. The Resonator reactor and the compensatory capacitor are decided by the lap of the circuit. The capacitive test sample can be replaced by the equivalent capacitance.

In order to improve the management level of the conductors to the withstand test better, a simple, friendly main interface is the first issue to be considered. Through the use of "Line", "Circle" and other functions and methods in the Form's Resize event, we have designed the main interface as shown in Figure 3. It is notable that the parameters of the functions cannot be the fixed value. Then, we use "Me.ScaleWidth * n" as the ones[5].



Fig. 3. The main interface of the system

3.4 The Design of Combining the Resonant Reactors

It is the core part of the system. According to the parameters and combinations of the reactor, the system should compute the total inductance, resistance and the allowed-maximum current and voltage automatically.

This section includes the table of available reactors, the table of the selected reactors and the lap preview. Lap preview can realize the function of visualization control to the resonant reactors. The design interface chart shows in Figure 4.

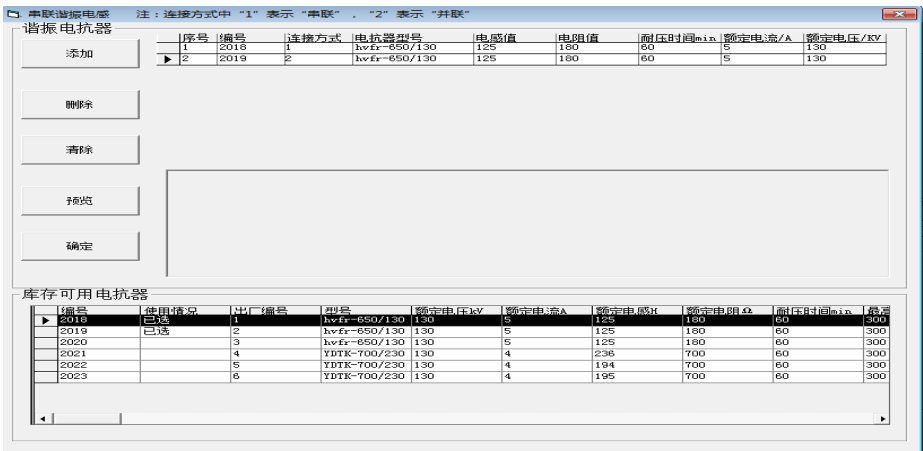


Fig. 4. The interface of combining the resonant reactors

In the lap of reactors, "1" indicates the reactor are in series, "2" indicates the reactor are in parallel. And series's priority is higher than the parallel's in computing. Firstly, judging whether the series connection or not, and then making a numerical tag. Finally,

judging the parallel connection and count the number of branch. The calculation logic diagram shows in Figure 5.

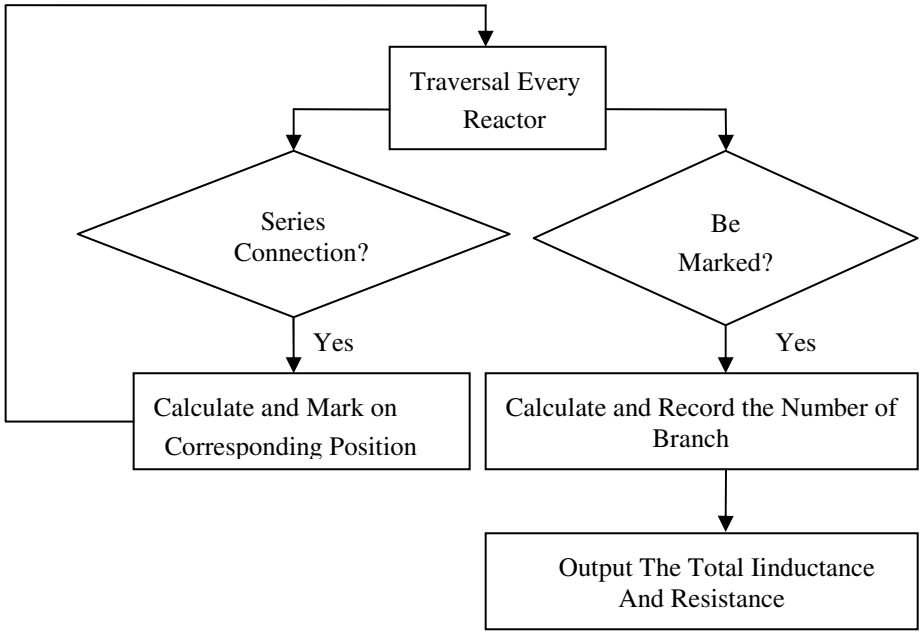


Fig. 5. The logic diagram of the Reactor parameters' calculation

4 Simulation

Following is a simulation example of a 110kV cable withstand voltage test provided by a power supply company.

4.1 The Parameters and the Plan of the Test

1) The parameters of the cable are as follows.

Table 1. Parameters of Cable

<i>Type</i>	<i>Length</i>	<i>Design of Capacitance</i>	<i>Test Voltage</i>
YJLW03 – 64/110kV	5.47 km	0.146 μ F/km	128kV

2) The selection and connection of the resonant reactors are as follows.

Table 2. Parameters of Reactors

Rating Voltage	Rating Currunt	Inductance	Resistance
130kV	5A	125H	180Ω
130kV	5A	125H	180Ω
130kV	5A	125H	180Ω
230kV	4A	238.81H	700Ω
230kV	4A	193.87H	700Ω
230kV	4A	193.12H	700Ω

*Note:*The relationship of the reactors is parallel

4.2 The Simulation of the Test

1) The manual lap of resonant reactors

Select 6 reactors from the the test equipment Library.And the combination between them is “2”, that is parallel. Click the "Preview" to see the connection of the reactors visually.The screenshot shows in Figure 6.

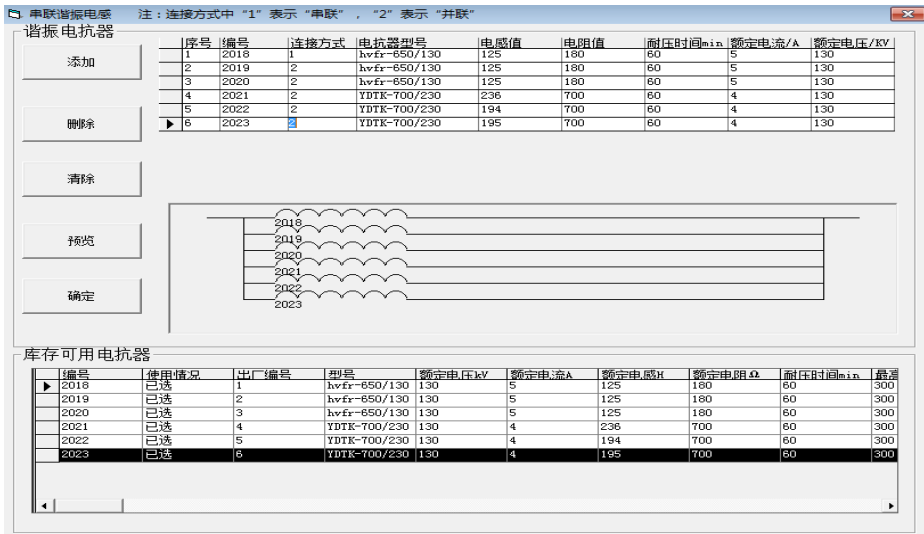


Fig. 6. The screenshot of the manual lap of resonant reactors

2) The result of the simulation

After inputting all kinds of electrical device’s parameters and running,the parameters needed should show on the circuit.After saving the test record,a test report is generated as WORD form dynamically.The screenshot of the result shows in Figure 7.

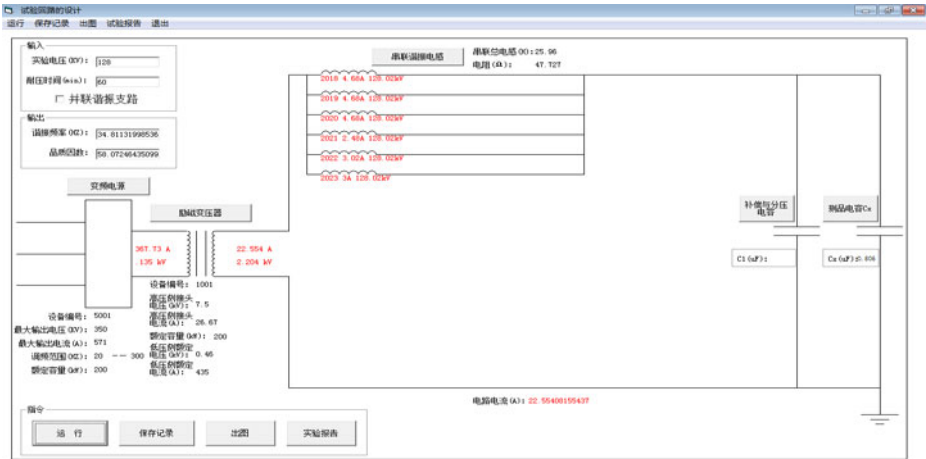


Fig. 7. The screenshot of the result

5 Conclusion

The simulation system of the withstand test has a very friendly interface ,simple operation and is easy to use. The site personnels can master it After a simple training.The system also has the function of saving the test record and displaying the previous test record.This will help the conductors manage the withstand test better.Because its design platform is very intuitive,it will be more conducive to maintenance of the system later.

The design principle of the system is entirely based on theoretical calculation, the withstand test in practice may also consider some environmental factors,such as the corona impacting on the quality factor. These questions will remain to be further research work.

References

1. Shi, F.: The Method of the XLPE Cable Withstand Test, Northeast Electric Power Technology, Beijing, China, pp. 47–49 (2010)
2. Li, W., Tu, Z.: Electrical Insulation Test and Measurement Equipment, pp. 53–57. China Electric Power Press, Beijing (2006)
3. Yao, W.: Visual Basic Database Development from Novice to Professional, pp. 124–126. Posts & Telecom Press, Beijing (2006)
4. Luo, J.: Research of Electrical Equipment Series Resonant Withstand Test Technology, Master thesis of South China University of Technology (2004)
5. Peng, B.: Visual Basic Programming Keys, pp. 60–62. Tsinghua University Press, Beijing (2008)

Improving Depth-First Search Algorithm of VLSI Wire Routing with Pruning and Iterative Deepening*

Xinguo Deng, Yangguang Yao, and Jiarui Chen

College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350108, China
Center for Discrete Mathematics and Theoretical Computer Science, Fuzhou University,
Fuzhou 350003, China

Abstract. A depth-first search (DFS) algorithm requires much less memory than breadth-first search (BFS) one. However, the former doesn't guarantee to find the shortest path in the VLSI (Very Large Integration Circuits) wire routing when the latter does. To remedy the shortcoming of DFS, this paper attempts to improve the DFS algorithm for VLSI wire routing by introducing a method of pruning and iterative deepening. This method guarantees to find all of the existing shortest paths with the same length in the VLSI wire routing to provide the wire routing designers more options for optimal designs.

Keywords: Depth-first search, Pruning, Iterative deepening, Wire routing, Shortest paths.

1 Introduction

Wire routing is a computation intensive task in the physical design of integrated circuits. With increasing chip sizes and a proportionate increase in circuit densities, the number of nets on a chip has increased tremendously. Typically, during the physical design of a VLSI (Very Large Integration Circuits) chip, it almost becomes mandatory to run the routing algorithm repeatedly in search of an optimal solution. [1]

Traditionally, routing is done in two stages of global routing and detailed routing sequentially [2, 3]. In the two stage routers, the global router abstracts the details of the routing architecture and performs routing on a coarser architecture. Then, the detailed router refines the routing done by the global router in each channel.

Lee [4] introduced an algorithm for routing a two terminal net on a grid in 1961. Since then, the basic algorithm has been improved for both speed and memory requirements. Lee's algorithm and its various improved versions form the class of maze routing algorithms.[5]

As noted in [6], the algorithm of *Rat in a Maze* does not guarantee to find a shortest path from maze entrance to exit. However, the problem of finding a shortest

* The work was supported by the Natural Science Foundation of Fujian Province (No.2009J05142), the Talents Foundation (No.0220826788) and the Scientific & Technological Development Foundation (No.2011-xq-24) of Fuzhou University.

path in a maze arises in the VLSI wire routing, too. To minimize signal delay, we wish to route the wire through a shortest path.

The algorithm of *Rat in a Maze* in [6] is essentially a DFS algorithm. To overcome the shortcoming of DFS algorithm, this paper focuses on improving the DFS algorithm for VLSI wire routing so as to achieve better wiring quality by simplifying the paths and therefore to reduce manufacturing costs and to increase the reliability. Our objective is to find all of the existing different shortest paths with the same length.

2 DFS Algorithm Improvement

2.1 Weakness of DFS Algorithm

The first path to the end traversed by a BFS algorithm is always the shortest one. A DFS algorithm requires much less memory than BFS one. However, the former doesn't guarantee to find the shortest path in the VLSI wire routing but the latter does.

Assuming figure 1 is a searching tree after optimizing the depth (A is the start position and F is the end position). There are various paths from the start to the end. Among them there is only one shortest path. The following are all possible cases: $A \rightarrow B \rightarrow D$ (not a solution), $A \rightarrow B \rightarrow E \rightarrow F$ (not an optimal solution), $A \rightarrow C \rightarrow E \rightarrow F$ (not an optimal solution), $A \rightarrow C \rightarrow F$ (the optimal solution). To overcome the blindness of DFS algorithm, the constraints of pruning and iterative deepening are added.

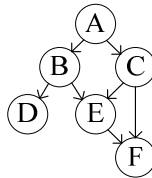


Fig. 1. An example of DFS

2.2 Pruning and Iterative Deepening

Pruning is a technique in machine learning that reduces the size of decision trees by removing sections of the tree that provide little power to classify instances. The dual goals of pruning are reduced complexity of the final classifier as well as better predictive accuracy by the reduction of over fitting and removal of sections of a classifier that may be based on noisy or erroneous data. [7]

Iterative deepening depth-first search (IDDFS) is a state space search strategy in which a depth-limited search is run repeatedly, increasing the depth limit with each iteration until it reaches d , the depth of the shallowest goal state. On each iteration, IDDFS visits the nodes in the search tree in the same order as DFS, but the cumulative order in which nodes are first visited, assuming no pruning, is effectively breadth-first. [8,9]

2.3 Measures Taken

Pruning can be used to improving the DFS algorithm so that the wire routing can be implemented with great efficiency and speed.

1. Variable *minStep* represents the shortest distance from a start to the end temporarily. *minStep* is initialized as a very big integer and dynamically changes in the course of search among various existing paths with different lengths between a start and the end. Variable *dep* represents the distance from a start to the current grid. If $dep \geq minStep$ and the current grid is not the end, the search should backtrack rather than proceed. The reason is that the current path is not one of the shortest paths in this case.

2. $dis[x][y]$ is a two dimensional array. $dis[x][y]$ represents the shortest distance from the start to the present grid temporarily. (nx, ny) is the neighbor of (x, y) . If $dis[x][y] + 1 > dis[nx][ny]$, the search should also backtrack rather than proceed. The reason is the same as in the case 1 above.

3. $matDis[x][y]$ is a two dimensional array. $matDis[x][y]$ represents the Manhattan distance [10] from the current grid (x, y) to the end (ex, ey) , i.e. $|x - ex| + |y - ey|$. It is the shortest distance between these two grids if there is no obstacle. If $dep + matDis[x][y] > minStep$, the search should backtrack rather than proceed too. The reason is that the distance of the current path is greater than the previous minimum distance *minStep*.

Pruning can dramatically shorten the search time in an ordinary situation. However, the longest distance from a start to the end is $n^2 - 1$ in a matrix $n \times n$. The large depth leads to the great search space. Even if the above pruning method is taken, the time complexity is still very large in the worst case.

Iterative deepening works by running depth-first search repeatedly with a growing constraint on how deep to explore the tree. This gives a search that is effectively breadth-first with the low memory requirements of DFS.

Except for pruning, constraint can be added to the distance *dep* between a start and the current grid. The increment of the distance *dep* with an initial value is one in a cyclic search. What's the range of the distance *dep*? It's between the Manhattan distance $matDis[sx][sy]$ and the longest distance from a start to the end. The repetition can be broken in advance if all shortest paths with the same length are found.

3 C++ Implementation

3.1 Design

The methodology of top-down modular is adopted to design the program. There are three basic aspects on the problem: input the maze, find all paths, and output all paths. A fourth module "Welcome" that displays the function of the program is also desirable. While this module is not directly related to the problem at hand, the use of such a module enhances the user-friendliness of the program. A fifth module "CalculateMemory" that calculates the memory is necessary here.

3.2 Program Plan

The design phase has already pointed out the need for five program modules. A root (or main) module invokes five modules in the following sequence: welcome module, input module, find all paths module, output module and calculate memory module.

A C++ program is designed by following the modular structure in Figure 2. Each program module is coded as a function. The root module is coded as the function “main”; “Welcome”, “InputMaze”, “FindAllPaths”, “DFS”, “CheckBound”, “ShowAllPaths”, “ShowOnePath” and “CalculateMemory” modules are implemented through different functions.

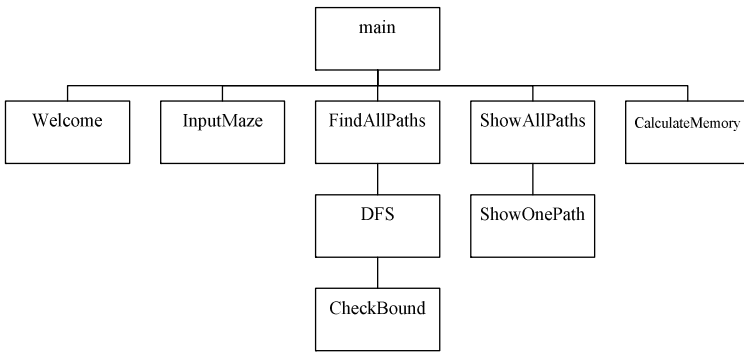


Fig. 2. Modular structure

3.3 Program Development

Function “Welcome” explains the function of the whole C++ program. Function “InputMaze” informs the user that the input is expected as a matrix of “0”s and “#”s besides a start and the end. The size of the matrix is determined first, so the number of rows and the number of columns of a matrix are needed before an input begins. It needs to be determined whether the matrix is to be provided by rows or by columns. In our experiment, the matrix is inputted by rows, and the input process is implemented by importing the input data from a text file called “in.txt”.

The idea of IDDFS is embodied in the function “FindAllPaths”. At first, the bool flag found is initialized with false. Then a cycle embedded the recursive subroutine DFS begins. The cyclic variable i with an increment of one represents the depth limited by IDDFS. It is between the Manhattan distance $matDis[sx][sy]$ and the longest distance from a start to the end. The repetition can be terminated ahead when all shortest paths with the same length are found.

The following figure 3 details the recursive function “DFS”. The four function parameters represent sequentially the horizontal coordinate, & the vertical coordinate, the distance from the start to the current grid, and the depth limited by IDDFS. The function returns in the following three conditions of pruning. 1. The depth of the current position is greater than the depth controlled; 2. The depth of the current position is greater than the previous minimum length when traversing from the start to the end; 3. The sum of the depth and the Manhattan distance of the current grid is

larger than the depth controlled. Otherwise, the minimum depth of the current grid is updated dynamically.

If the present position is the end, then a path is found. The bool flag *found* is changed to true. The current depth is exactly the shortest distance from a start to the end. To construct a shortest path between the start and the end, traversal begins from the end to the start. The movement is from the present position to its neighbor labeled one less. Such a neighbor must exist as each grid's label is one more than that of at least one of its neighbors. While back to start, coordinates of a shortest path are reserved in a vector. The vector is added to another two dimensional vector representing all shortest paths. The current recursive function returns to the upper one.

If the present position is not the end, its east, south, west or north neighbor is checked sequentially. If the neighbor is within the bound and is not blocked and the depth of the current position after one increment is not larger than the depth of the neighbor, the next step proceeds. The visited label of the neighbor is marked. Coordinates of the neighbor are reserved in the matrix of the path. The recursive function "DFS" is called after the depth of the current position is increased by one. This means that the recursive search continues from the neighbor. The label of visited the neighbor is cancelled before the traversal backtracks.

The function "Checkbound" is used to judge whether a position is within the reasonable bound or not. It is called by the recursive function "DFS". The details of the functions "ShowAllPaths", "ShowOnePath" and "CalculateMemory" are omitted here. The effect of these functions will be illustrated in next section.

```

//finds all shortest paths
//coordinates,current depth,depth controlled by IDDFS
void Circuit::DFS(int x,int y,int dep,int curDep)
{
    if(dep<minStep || dep>curDep) //pruning 1 & 2
        return ;
    if(matDis[x][y]+dep>curDep) //pruning 3
        return ;
    dis[x][y]=dep; //minimum depth
    if(mat[x][y]=='E')
    {
        found=true;
        minStep=curDep;

        vector<ipair> OnePath; //vector of one shortest path
        for(int i=dep-1 ;i>=0 ;i--)
            OnePath.push_back(res[i]);
        OnePath.push_back(mp(sx,sy));

        path.push_back(OnePath); //all shortest paths
        return ;
    }
    for(int i=0 ;i<4 ;i++) //possible direction
    {
        int nx=x+dx[i];
        int ny=y+dy[i];
        //check bound,not visited,dep+1<=dis[nx][ny]
        if(CheckBound(nx,ny) && !visited[nx][ny] && dep+1<=dis[nx][ny])
        {
            visited[nx][ny]=true; //visited
            res[dep]=mp(nx,ny); //reserve coordinates
            DFS(nx,ny,dep+1,curDep); //recursive calling
            visited[nx][ny]=false; //cancell label visited before backtacking
        }
    }
}
}

```

Fig. 3. The recursive function DFS

4 Algorithm Complexity and Experimental Results

4.1 Algorithm Complexity

For analysis of the algorithm complexity, the worst case is introduced when there is no obstacle in the maze at all. From interior (i.e. non-boundary) positions of the maze, four moves are possible: east, south, west or north. From positions on the boundary of the maze, either two or three moves are possible. However, if the present position is not the end, it is marked "visited" so as to prevent the search from returning here. Thus, the count of moves should be one less. The upper bound for the recursive function "DFS" is $O(a^{2n})$ for a maze of $n \times n$ matrix theoretically ($2 < a < 3$). The actual upper bound should be less than $O(a^{2n})$ owing to the limit of pruning and iterative deepening. The more obstacles there are, the less the complexity of the recursive function "DFS" is. The reason is that the count of moves decreases with the increase of obstacles.

4.2 Experimental Results

Figure 4 is partial data of the output about a random maze of 17×17 matrix. The figure shows that the length of shortest paths is 38 and there are 160 different paths from the start to the end in all. Then the coordinates and the map of these paths are output respectively. Only one of them is shown below. The data of other 159 paths are omitted here due to the limit of the paper. The next is the final map of these 160 paths overlapped. The last are the total running time and the memory occupied.

The amount of the shortest paths with the same length, running time and memory consumed of the program vary greatly with the size of the maze and the distribution of obstacles. The worst case occurs when there is not any obstacle in the maze at all. Table 1 displays part of the statistic data of the worst case. The table shows that paths, time and memory rise sharply along with the increase of the size of the maze.

Table 1. Partial statistic data of worst case

n	paths	time(ms)	Memory (kB)	a^{2n} ($a=2$)
2	2	0	326	16
3	6	0	326	64
4	20	15	327	256
5	70	15	330	1024
6	252	62	345	4096
7	924	218	412	16384
8	3432	734	701	65536
9	12870	2859	1934	262144
10	48620	11734	7163	1048576
11	184756	49625	29194	4194304
12	705432	211391	121572	16777216

```

////////////////////////////////////
This program uses the improved Depth-First-Search algorithm to
find all shortest paths from the start point to the end point.
////////////////////////////////////

Enter the row size of gird:
Enter the column of gird:
Enter wiring gird
S :the start point
E :the end point
0 :the place we can traverse
# :the obstacle
The length of shortest path(s) is 38.
There are 160 path(s) from 'S' to 'E'
path 1 (coordinates):
0 0
0 1
0 2
0 3
0 4
0 5
0 6
1 6
1 7
1 8
0 8
0 9
0 10
1 10
1 11
2 11
2 12
3 12
3 13
4 13
5 13
6 13
7 13
7 14
8 14
8 15
9 15
10 15
10 16
11 16
12 16
13 16
13 15
13 14
14 14
15 14
15 15
15 16
16 16

path 1 (map):
S  1  2  3  4  5  6  #  10 11 12 0 0 #  0 0 #
0  0  0  0  0  0  #  7  8  9  #  13 14 0 #  0 0 #
0  0  #  0  0  0  #  #  #  #  0  #  15 16 0 0 #  0
0  0  0  0  #  0  0  0  0  0  0  #  17 18 0 0 #  0
0  0  0  0  0  #  0  0  #  #  #  #  0  0  #  19 #  0  0
0  #  #  #  #  #  #  0  0  #  0  0  #  0  20 0  #  0
0  #  0  0  0  #  0  0  0  0  #  0  #  0  21 0  0  #
0  #  0  0  0  0  #  #  0  0  0  #  #  0  22 23 0  0
0  0  0  0  0  0  0  0  #  0  0  #  #  0  #  24 25 0
0  #  #  #  #  #  #  #  #  #  #  0  0  0  0  #  26 #
0  0  0  #  0  0  0  0  0  0  #  0  0  #  0  27 28
0  0  0  #  0  0  #  #  #  #  #  0  0  0  0  #  29
0  #  #  #  0  0  0  0  0  0  0  #  0  #  0  30
0  0  #  0  0  0  0  0  0  #  #  0  0  #  33 32 31
0  #  0  0  0  0  0  #  #  0  0  #  0  34 # #
0  0  0  0  #  0  0  #  0  0  0  0  #  0  35 36 37
0  0  0  0  #  0  0  #  0  0  0  0  0  0  #  E

... ..

The final map :
S  1  2  3  4  5  6  #  10 11 12 13 14 #  0 0 #
0  0  0  0  0  0  #  7  8  9  #  13 14 15 #  0 0 #
0  0  #  0  0  0  #  #  #  #  0  #  15 16 17 0 #  0
0  0  0  #  0  0  0  0  0  0  0  0  #  17 18 0 0 #  0
0  0  0  0  0  #  0  0  #  #  #  #  0  0  #  19 #  0  0
0  #  #  #  #  #  #  0  0  #  0  0  #  0  20 21 #  0
0  #  0  0  0  #  0  0  0  0  #  0  #  0  21 22 23 #
0  #  0  0  0  0  #  #  0  0  0  #  #  0  22 23 24 0
0  0  0  0  0  0  0  0  #  0  0  #  #  0  #  24 25 0
0  #  #  #  #  #  #  #  #  #  #  0  0  0  0  #  26 #
0  0  0  #  0  0  0  0  0  0  #  0  0  #  0  27 28
0  0  0  #  0  0  #  #  #  #  #  0  0  0  0  #  29
0  #  #  #  0  0  0  0  0  0  0  #  0  #  0  31 30
0  0  #  0  0  0  0  0  0  #  #  0  0  #  33 32 31
0  #  0  0  0  0  0  #  #  0  0  #  0  34 # #
0  0  0  0  #  0  0  #  0  0  0  0  #  0  35 36 37
0  0  0  0  #  0  0  #  0  0  0  0  0  0  #  E

The total running time is 141 ms
The total memory is 373 KB

```

Fig. 4. Partial output data about a random maze of 17x17 matrix

5 Conclusion

In order to find all shortest paths with the same length in the VLSI wire routing, the depth-first search algorithm is improved with the method of pruning and iterative deepening, providing wire routing designers various options to optimize their designs. Therefore, a C++ program is developed to implement the enhancing DFS algorithm. Further, the satisfactory experimental results of running the C++ program are presented. This new method guarantees to find all existing shortest paths with the same length in the VLSI wire routing with only moderate computer memory consumption.

Nevertheless, more refinement work of the aforementioned enhanced DFS algorithm is needed. The upper bound of the algorithm complexity is $O(a^{2n})$ for a maze of $n \times n$ matrix theoretically ($2 < a < 3$). Even if experimental results show that the actual upper bound is far less than $O(a^{2n})$, the algorithm needs further better before its practical application. Future work includes the improvement of the efficiency of the algorithm and further reducing the computer memory consumption.

References

1. Kumar, H., Kalyan, R., Bayoumi, M., Tyagi, A., Ling, N.: Parallel implementation of a cut and paste maze routing algorithm. In: Proceedings of IEEE International Symposium on Circuits and Systems ISCAS 1993, vol. 3, pp. 2035–2038 (1993)
2. Taghavi, T., Ghiasi, S., Sarrafzadeh, M.: Routing algorithms: architecture driven rerouting enhancement for FPGAs. In: Proceedings of IEEE International Symposium on Circuits and Systems ISCAS 2006, pp. 5443–5446 (2006)
3. Wolf, W.: Modern VLSI Design: System-on-Chip Design, 3rd edn., pp. 518–522. Pearson Education, Inc, London (2003)
4. Lee, C.Y.: An algorithm for path connections and its applications. IRE Trans. Electronic Computers (September 1961)
5. Sherwani, N.A.: Algorithms for VLSI Physical Design Automation, 3rd edn., pp. 286–288. Kluwer Academic Publishers, Dordrecht (2002)
6. Sahni, S.: Data Structures, Algorithms, and Applications in C++, 2nd edn., pp. 268–279. McGraw-Hill, New York (2004)
7. Kantardzic, M.: Data Mining: Concepts, Models, Methods, and Algorithms, 1st edn., pp. 139–164. Wiley-IEEE Press (2002)
8. Ibrahim, A., Fahmi, S.A., Hashmi, S.I., Ho-Jin, C.: Addressing Effective Hidden Web Search Using Iterative Deepening Search and Graph Theory. In: Proceedings of IEEE 8th International Workshops on Computer and Information Technology, CIT 2008, pp. 145–149 (July 2008)
9. Cazenave, T.: Optimizations of data structures, heuristics and algorithms for path-finding on maps. In: Proceedings of IEEE International Symposium on Computational Intelligence and Games, CIG 2006, pp. 27–33 (2006)
10. Dar-Jen, C., Desoky, A.H., Ming, O.Y., Rouchka, E.C.: Compute Pairwise Manhattan Distance and Pearson Correlation Coefficient of Data Points with GPU. In: Proceedings of IEEE International Conference on Software Engineering, Artificial Intelligences, Networking and Parallel/Distributed Computing, SNPD 2009, pp. 501–506 (2009)

The Forecast of Our Private Car Ownership with GM(1,1) Model

Yi Zhang

College of Mathematics and Information,
China West Normal University; Nanchong, Sichuan, 637002
zhangyi3181w@163.com

Abstract. The data of our private car ownership in recent years has been analyzed, a strong trend of exponential growth has been found in this data, and the GM (1,1) model for this data has been established in this paper. The result is given to show that the precision of the model is high, and is used to forecast the future of the private car ownership.

Keywords: Car, Grey GM (1,1) model, forecast.

1 Introduction

With the continuous development of China's economy and per capita income levels rising, and national implementation of the "car to the countryside" and other related incentives for the auto consumption policy, car is no longer a distant luxury, but gradually into the ordinary family, became the transport of ordinary people, private car ownership in China has entered a high growth era. But with the rapid growth of private car ownership, solving the urban construction, environmental protection, traffic congestion and other problems become more difficult, in order to make these issues are more properly addressed, we need analyze the forecast of private car ownership, so as to provide more accurate data for countries to develop policies and measures.

According to the data on China's private car ownership in 1995-2007, with time as the X-axis, with the data of the national private car ownership as Y-axis, we making the curve, from which we can see that the data of private car ownership increased with the growth of time, and showed a certain trend of exponential growth. Therefore, this paper argues that we can establish a model with the data of private car ownership, and give the prediction by the model. Because of the small amount and the trend of exponential growth of the data, using the traditional statistical methods are difficult to analyze and research. The GM (1,1) model proposed by Mr. Deng Julong to deal with the data which has the features of "less samples, poor information", is usually able to obtain high accuracy. Thus, we use GM (1,1) model to simulate them.

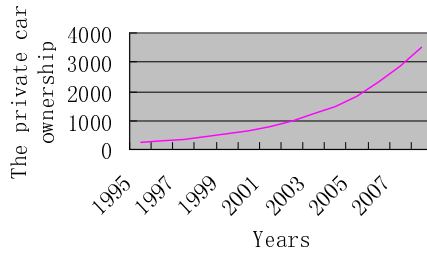


Fig. 1. The data of China's private car ownership in 1995-2007

2 The GM(1,1) Model

The GM(1,1) model applied in many areas since it was introduced by professor Deng. As this model needs fewer samples and the computation is simple, it has superiority than the traditional estimate method.

Theorem: Let $x^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\}$ be nonnegative pre-smooth series and $x^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)\}$ be the AGO series of $x^{(0)}$. We call $x^{(0)}(k) + az^{(1)}(k) = b \quad k = 2,3,\dots$ grey differential equation, and $\frac{dx^{(1)}}{dt} + ax^{(1)} = b$ whitened differential equation. $\alpha = (a,b)^T$ is the parameter of the equation.

$$\text{Let } Y = \begin{pmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{pmatrix} \text{ and } B = \begin{pmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{pmatrix}$$

Then, 1) From to the method of least square, we have $\hat{\alpha} = (B^T B)^{-1} B^T Y$;

2)The continuous solution of the equation is

$$x^{(1)}(k) = \left(x^{(1)}(1) - \frac{b}{a} \right) e^{-a(k-1)} + \frac{b}{a};$$

3) The discrete solution of the equation

$$\text{is } x^{(1)}(k) = \left(x^{(1)}(1) - \frac{b}{a} \right) e^{-a(k-1)} + \frac{b}{a} \quad k = 2,3,\dots;$$

4)The original value is $\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \quad k = 1,2,\dots$

3 The Forecast of Our Private Car Ownership

The data of private car ownership from the "National Statistical Yearbook" is as follows. (Unit: million volume).

Table 1. The data of China's private car ownership in 1995-2007

Years	1995	1996	1997	1998	1999	2000	2001
data	249.96	289.67	358.36	423.65	533.88	625.33	770.78
Years	2002	2003	2004	2005	2006	2007	2008
data	968.98	1219.23	1481.66	1848.07	2333.32	2876.22	3501.39

We use GM(1,1)model to simulate the data from 1995 to 2005 by MATLAB. And the model is as follow.

$$x^{(1)}(k) = 1172.78e^{0.2095(k-1)} - 9228.47 \quad k = 2,3,\dots \quad (1)$$

The simulated value and relative error is in the following table.

Table 2. The simulated value and relative error

years	Car ownership	Simulated value	relative error (%)	years	Car ownership	Simulated value	relative error (%)
1996	289.67	273.33	5.64	2003	1219.23	1184.56	2.84
1997	358.36	337.02	5.95	2004	1481.66	1460.63	1.42
1998	423.65	415.57	1.91	2005	1848.07	1801.04	2.54
1999	533.88	512.42	4.02	2006	2333.32	2220.79	4.82
2000	625.33	631.84	-1.04	2007	2876.22	2738.36	4.79
2001	770.78	779.10	-1.08	2008	3501.39	3376.54	3.57
2002	968.98	960.67	0.86				

From the data of tables 2, we have that: the simulation precision of the GM (1, 1) model is high, which is higher than 94%. We predict the ownership from 2006 to 2008 by the model, and we can see the prediction precision is also high, which is more than 95%. Thus, the model is effective and feasible.

In order to obtain more accurate forecasting results, following the Grey's "new information first " principle, we use the data from 1999 to 2008 to model a new GM(1,1) model as follow:

$$x^{(1)}(k) = 2653.86e^{0.2140(k-1)} - 2119.98 \quad k = 2,3,\dots \tag{2}$$

The simulated value and relative error is in the following table.

Table 3. The simulated value and relative error

years	Car ownership	Simulated value	relative error (%)	years	Car ownership	Simulated value	relative error (%)
2000	625.33	633.22	-1.26	2005	1848.07	1845.94	0.12
2001	770.78	784.31	-1.76	2006	2333.32	2286.39	2.01
2002	968.98	971.45	-0.26	2007	2876.22	2831.93	1.54
2003	1219.23	1203.24	1.31	2008	3501.39	3507.64	-0.18
2004	1481.66	1490.34	-0.59				

Thus, the simulated error of the new model according to the "new information first " principle is reduced greatly and the simulation precision has reached 97%. Therefore, we adopt this model of private car ownership in China were forecast to reach better results. According to this model , we get the result of prediction: the ownership in 2010 is 5381.22 million volume; and in 2015 is 15687.1 million volume. This shows that China's private car ownership will have a fast growth in the coming years.

4 Conclusions

Through the analysis of private car ownership, we give a GM(1,1) model and its prediction. Of course, this prediction is mainly based on ownership's own rule in the development and changes, but did not consider the various factors affecting the automobile industry, such as: GDP, consumer environment, energy supply, but the overall sense, with the sustained health of our country's development and continuous improvement of people's consumption level, China's private car ownership will maintain rapid growth trend in the next longer period.

References

1. Deng, J.: Estimate and Decision of Grey System. Huazhong University of Science & Technology Press, Wuhan (2002) (in Chinese)
2. Zhang, Y., Wei, Y., Xiong, C.-w.: One new optimized method of GM (1, 1) model. *Systems Engineering Theory & Practice* (4), 141–146 (2007)
3. Zhang, Y., Wei, Y., Zhou, P.: The Improved Approach of Grey Derivative in GM(1,1) Model. *The Journal of Grey System* (4), 375–380 (2006)
4. Zhang, Y., Wei, Y., Xiong, C.-w.: An Approach of GM(1,1) Based on Optimum Grey Derivative. *The Journal of Grey System(UK)* (4), 397–404 (2007)
5. Zhang, Y., Wei, Y., Xiong, C.-w.: A New Method to Find The Original Value of Various GM(1,1) With Optimum Background Value. In: 2008 IEEE Intl Conference on Systems, Man and Cybernetics, vol. 10, pp. 2476–2480 (2008)
6. National Bureau of Statistics, China Statistical Yearbook. China Statistics Press, Beijing (2009)
7. Society of Automotive Engineers, Automotive Blue Book. Social Sciences Academic Press, Beijing (2008)

The Rationality Properties of Three Preference-Based Fuzzy Choice Functions*

Xue Na, Yixuan Ma** and Yonghua Hao

Department of Mathematics of TUT, Taiyuan 030024
xuena03@163.com

Abstract. In this paper, three $PFCF_s$ are selected and whether they satisfy the rationality properties such as $F\alpha$, $F\beta$, $F\gamma$, $WFCA$ is investigated, then the satisfaction of preference-based fuzzy choice functions is obtained as a consequence.

Keywords: Fuzzy Preference Relation, Preference-based Fuzzy Choice Function, Rationality Property.

1 Introduction

In the research on choice functions, there is lots of literature in which the relationships between rationality conditions and the characterization of the rationalization of choice functions are discussed whereas little attention is paid to the rationality of some specific choice functions. The rationality of some specific choice functions include: Barrett [1] proposed nine preference-based non-fuzzy choice functions and assess whether those satisfy $RPWD$, $RPSD$, $SREJ$, $WREJ$, UF , LF ; Roubens [2] defined three preference-based fuzzy choice functions and analyzed some properties of these choice functions. Based on them, in this paper we select three preference-based fuzzy choice functions and assess these preference-based fuzzy choice functions selected in terms of some usual rationality conditions in fuzzy case.

2 Preliminaries

Let X denote a finite set of alternatives, $P(X)$ the set of all non-empty crisp subsets of X , and $F(X)$ the set of all non-empty fuzzy subsets of X .

Definition 1 [1]: A fuzzy binary preference relation ($FBPR$) is a function

$$R : X \times X \rightarrow (0,1).$$

* The project is supported by the Natural Science Foundation of Shanxi Province (2009011006) and Yong Foundation of TUT.

** Corresponding author.

Remark 1. An *FBPR* R satisfies transitivity iff

$$\text{for all } x, y, z \in X, R(x, z) \geq R(x, y) \wedge R(y, z).$$

Remark 2. G denotes the non-empty set of all *FBPRs*.

Definition 2 [1]: A fuzzy choice function is a function

$$C : P(X) \rightarrow F(X) \text{ such that } \forall S \in P(X), \text{sup } pC(S) \subseteq S,$$

where $\text{sup } pS$ denotes the support of S for all $S \in F(X)$. We denote it as $C(\cdot)$.

Remark 3. A fuzzy choice function $C(\cdot)$ is called a preference-based fuzzy choice function (*PFCF*) iff $C(\cdot)$ is determined by a fuzzy relation Q . We denote it as $C(\cdot, Q)$.

The choice set $C(S)$ is normal iff $\forall S \in P(X), \exists x \in S$, such that $C(S) = 1$.

Definition 3 [3]: Let $C(\cdot)$ be a fuzzy choice function, the fuzzy preference relation revealed by $C(\cdot)$ is defined by

$$\forall x, y \in X, \forall S \in P(X), R(x, y) = \bigvee_{\{S|x, y \in S\}} C(S)(x).$$

Remark 4. It is clear that $\forall x \in X, R(x, y) \geq C(S)(x) (\forall y \in S)$.

In this section we select three *PFCFs* as follows:

Let Q is a *FBPR*, $\forall x, y \in X, \forall S \in P(X)$, then we denote

$$(1) C_1(S, Q)(x) = \bigwedge_{y \in S} Q(x, y).$$

$$(2) C_2(S, Q)(x) = 1 - \bigvee_{y \in S} P(y, x) \text{ such that } P(y, x) = (Q(y, x) - Q(x, y)) \vee 0.$$

Obviously $P(x, x) = 0$.

$$(3) C_3(S, Q)(x) = \bigvee_{y \in S} P(x, y).$$

Remark 5. In general, the three functions above may not be non-empty, so we cannot assure those are choice functions. The following results are based on the hypothesis that those three *PFCFs* are all choice functions.

Remark 6. $C_1(S, Q)(x)$ is the same as $D^Q(x)$ [2], it means the degree to which x is preferred to any element of S in terms of Q ; $C_2(S, Q)(x)$ is the same as $N^Q(x)$ [2], it means the degree to which any element of S is not preferred to x in terms of Q ; $C_3(S, Q)(x)$ is different from $SD^Q(x)$ [2], it means the degree to which x is strictly preferred to any element of S in terms of Q .

3 Some Rationality Properties of Preference-Based Choice Functions

In this section we consider some restrictions on *PFCFs*, which may be viewed as representing different aspects of rationality. Some of these properties have very well known counterparts in the vast literature [3,4,5] and we examine, in terms of these properties, the performance of the different *PFCFs* defined in section 2. Now we list these properties as follows:

$$F\alpha: \forall S_1, S_2 \in P(X), \forall x \in S_1, S_1 \subseteq S_2 \Rightarrow C(S_2)(x) \leq C(S_1)(x).$$

$$F\beta: \forall S_1, S_2 \in P(X), \forall x, y \in S_1,$$

$$S_1 \subseteq S_2 \Rightarrow C(S_1)(x) \wedge C(S_1)(y) \wedge C(S_2)(x) \leq C(S_2)(y).$$

$$F\gamma: \forall S_1, S_2 \in P(X), \forall x \in S_1 \cap S_2, C(S_1)(x) \wedge C(S_2)(x) \leq C(S_1 \cup S_2)(x).$$

$$F\beta^+ : \forall S_1, S_2 \in P(X), \forall x, y \in S_1, S_1 \subseteq S_2 \Rightarrow C(S_1)(x) \wedge C(S_2)(y) \leq C(S_2)(x) .$$

$$FA1: \forall A, B \in P(X), x \in A \cap B, \text{ then } C(A)(x) \wedge C(B)(x) = C(A \cup B)(x),$$

[5] point out : $FA1 \Leftrightarrow F\alpha + F\gamma$.

$$FA2: \forall A, B \in P(X), \text{ if } B \subseteq A - \text{sup } pC(A), \text{ then } C(A - B) \subseteq C(A).$$

Obviously, $FA2$ is satisfied iff $\forall x \in B, C(A)(x) = 0$ for $\forall A, B \in P(X)$,

$$\text{then } C(A - B)(x) \leq C(A)(x) \text{ for } \forall x \in A - B.$$

$$WFCA: \forall S \in P(X), \forall x, y \in S, R(x, y) \wedge C(S)(y) \leq C(S)(x).$$

$$FC1: \forall S \in P(X), \forall x, y \in S, \{y \text{ is dominant in } S\} \Rightarrow C(S)(x) = R(x, y).$$

$$FC2: \forall S \in P(X), \forall x, y \in S,$$

$$\{y \text{ is dominant in } S \text{ and } R(x, y) \geq R(y, x)\} \Rightarrow x \text{ is dominant in } S.$$

$$FC3: \forall S \in P(X), \forall x, y \in S \text{ and all real numbers } k \text{ such that } 0 < k \leq 1, \\ \{C(S)(y) \geq k \text{ and } R(x, y) \geq k\} \Rightarrow C(S)(x) \geq k.$$

4 Result

Proposition 1. (1) $C_1(\cdot, Q)$ and $C_2(\cdot, Q)$ satisfy $F\alpha$.

(2) $C_3(\cdot, Q)$ violates $F\alpha$.

Proof

(1) Let $\forall S_1, S_2 \in P(X), S_1 \subseteq S_2, S_1 \subseteq S_2, \forall Q \in G, \forall x \in S_1$

$$C_1(S_2, Q)(x) = \bigwedge_{z \in S_2} Q(x, z) \leq \bigwedge_{z \in S_1} Q(x, z) = C_1(S_1, Q)(x)$$

$$C_2(S_2, Q)(x) = 1 - \bigvee_{z \in S_2} P(z, x) \leq 1 - \bigvee_{z \in S_1} P(z, x) = C_2(S_1, Q)(x)$$

so $C_1(\cdot, Q)$ and $C_2(\cdot, Q)$ satisfy $F\alpha$.

(2) The proof consists of counterexamples.

$$\text{Let } X = \{x, y, z\}, S_1 = \{x, y\}, S_2 = X, Q = \begin{pmatrix} 1 & 0.6 & 0.6 \\ 0.2 & 1 & 0.2 \\ 0.1 & 0.3 & 1 \end{pmatrix}.$$

Then $C_3(S_2, Q)(x) > C_3(S_1, Q)(x)$,

But $S_1 \subseteq S_2$, so $C_3(\cdot, Q)$ violates $F\alpha$. □

Proposition 2. (1) $C_3(\cdot, Q)$ satisfies $F\beta^+$.

(2) $C_1(\cdot, Q)$ and $C_2(\cdot, Q)$ violate $F\beta$.

Proof

(1) Let $\forall S_1, S_2 \in P(X), \forall x, y \in S_1, S_1 \subseteq S_2, \forall Q \in G$,

$$C_3(S_1, Q)(x) \wedge C_3(S_2, Q)(y) = [\bigvee_{z \in S_1} P(x, z)] \wedge [\bigvee_{z \in S_2} P(y, z)]$$

$$\leq \bigvee_{z \in S_2} P(x, z) = C_3(S_2, Q)(x),$$

so $C_3(\cdot, Q)$ satisfies $F\beta^+$.

(2) Let $X = \{x, y, z\}, S_1 = \{x, y\}, S_2 = X$,

$$\text{Consider } Q_1, Q_2 \text{ such that } Q_1 = \begin{pmatrix} 1 & 0.6 & 0.4 \\ 0.8 & 1 & 0.3 \\ 0.6 & 0.3 & 1 \end{pmatrix}, Q = \begin{pmatrix} 1 & 0.8 & 0.2 \\ 0.6 & 1 & 0.3 \\ 0.4 & 0.6 & 1 \end{pmatrix}.$$

By computing, we have

$$C_1(\cdot, Q_1), C_2(\cdot, Q_2) \text{ violates } F\beta. \quad \square$$

Remark 7. (1) By $F\beta^+ \Rightarrow F\beta$, $C_3(\cdot, Q)$ satisfies $F\beta$.

(2) $F\beta^+$ is stronger than $F\beta$ so $C_1(\cdot, Q)$ and $C_2(\cdot, Q)$ violate $F\beta^+$.

Proposition 3. $C_1(\cdot, Q), C_2(\cdot, Q)$ and $C_3(\cdot, Q)$ all satisfy $F\gamma$.

Proof Let $\forall S_1, S_2 \in P(X), \forall x \in S_1 \cap S_2, \forall Q \in G$,

$$(1) C_1(S_1, Q)(x) \wedge C_1(S_2, Q)(x) = [\bigwedge_{z \in S_1} Q(x, z)] \wedge [\bigwedge_{z \in S_2} Q(x, z)]$$

$$= \bigwedge_{z \in S_1 \cup S_2} Q(x, z) = C_1(S_1 \cup S_2, Q)(x),$$

so $C_1(\cdot, Q)$ satisfies $F\gamma$.

$$(2) \bigvee_{z \in S_1 \cup S_2} P(z, x) = [\bigvee_{z \in S_1} P(z, x)] \wedge [\bigvee_{z \in S_2} P(z, x)]$$

$$\text{suppose } \bigvee_{z \in S_1} P(z, x) \geq \bigvee_{z \in S_2} P(z, x)$$

$$\text{then } C_2(S_1, Q)(x) \wedge C_2(S_2, Q)(x) = [1 - \bigvee_{z \in S_1} P(z, x)] \wedge [1 - \bigvee_{z \in S_2} P(z, x)]$$

$$= [1 - \bigvee_{z \in S_1} P(z, x)] = [1 - \bigvee_{z \in S_1 \cup S_2} P(z, x)] = C_2(S_1 \cup S_2, Q)(x)$$

so $C_2(\cdot, Q)$ satisfies $F\gamma$.

$$(3) C_3(S_1, Q)(x) \wedge C_3(S_2, Q)(x) = [\bigvee_{z \in S_1} P(x, z)] \wedge [\bigvee_{z \in S_2} P(x, z)]$$

$$\leq \bigvee_{z \in S_1} P(x, z) \leq \bigvee_{z \in S_1 \cup S_2} P(x, z) = C_3(S_1 \cup S_2, Q)(x)$$

so $C_3(\cdot, Q)$ satisfies $F\gamma$. □

By [5], $FA1 \Leftrightarrow F\alpha + F\gamma$, then the following proposition is satisfied.

Proposition 4. $C_1(\cdot, Q)$ and $C_2(\cdot, Q)$ satisfy $FA1$.

Remark 8. By $C_3(\cdot, Q)$ violates $F\alpha$, so it violates $FA1$ (See the proof of proposition(2)).

Proposition 5. (1) $C_3(\cdot, Q)$ satisfies $FA2$.

(2) $C_1(\cdot, Q)$ and $C_2(\cdot, Q)$ violate $FA2$.

Proof

(1) Let $\forall A, B \in P(X), B \subseteq A - \sup pC(A), \forall x \in A - B, \forall Q \in G$,

$$\text{then } C_3(A - B, Q)(x) = \bigvee_{z \in A - B} P(x, z) \leq \bigvee_{z \in A} P(x, z) = C_3(A, Q)(x),$$

so $FA2$ is satisfied.

(2) We only proof $C_2(\cdot, Q)$ violate $FA2$, the proof of $C_1(\cdot, Q)$ is similar.

$$\text{Let } X = \{x, y, z\}, A = \{x, y, z\}, B = \{y\}, Q = \begin{pmatrix} 1 & 1 & 0.8 \\ 0 & 1 & 0.9 \\ 0.2 & 0.2 & 1 \end{pmatrix}.$$

then $A - \sup pC_2(A, Q) = \{y\}, B \subseteq A - \sup pC_2(A, Q)$,

But $C_2(A - B, Q)(z) > C_2(A, Q)(z)$,

so $C_2(\cdot, Q)$ violates $FA2$. □

Proposition 6. (1) If Q satisfies transitivity then $C_1(\cdot, Q)$ satisfies $WFCA$.

(2) $C_2(\cdot, Q), C_3(\cdot, Q)$ violate $WFCA$.

Proof

(1) Let $\forall S \in P(X), \forall x, y \in S, \forall Q \in G$ and Q satisfies transitivity,

$$R(x, y) \wedge C_1(S, Q)(y) = Q(x, y) \wedge [\bigwedge_{z \in S} Q(y, z)] = \bigwedge_{z \in S} [Q(x, y) \wedge Q(y, z)]$$

$$\leq \bigwedge_{z \in S} Q(x, z) = C_1(S, Q)(x)$$

then $WFCA$ is satisfied.

(2) We only proof $C_2(\cdot, Q)$ violate $WFCA$, the proof of $C_3(\cdot, Q)$ is similar.

Let $X = \{x, y, z\}$, $Q = \begin{pmatrix} 1 & 0.3 & 0.1 \\ 0.8 & 1 & 0.3 \\ 0.8 & 0.7 & 1 \end{pmatrix}$, then $C_2(\cdot, Q)$ violates $WFCA$. □

In the literature [6], Guo proposed $WFCA \Leftrightarrow FC3 \Leftrightarrow F\alpha + F\beta$ in the hypothesis that $T = \wedge$ and the choice set is normal; by [7], we have $FC3 \Rightarrow FC2$, so the following notation is hold

Remark 9. If the choice set is normal and Q satisfies transitivity then $C_1(\cdot, Q)$ satisfies $WFCA$, $FC3$ and $FC2$.

5 Conclusions

In this paper, we have investigated the rationality properties of three preference-based choice functions, our main results are summarized in the following table.

Table 1. The rationality of preference-based fuzzy choice functions

	$C_1(\cdot, Q)$	$C_2(\cdot, Q)$	$C_3(\cdot, Q)$
$F\alpha$	*	*	
$F\beta$	*"		*
$F\beta^+$			*
$F\gamma$	*	*	*
$FA1$	*	*	
$FA2$			*
$WFCA$	*		

In the table, * indicates that the $PFCF$ satisfies the property under consideration; *' indicates that when Q satisfies transitivity the $PFCF$ satisfies the property; *" indicates that the $PFCF$ satisfies the property under the condition that the choice set is normal and Q satisfies transitivity.

Remark 10. In addition, $C_1(\cdot, Q)$ satisfies $FC3$ and $FC2$ under the condition that the choice set is normal and Q satisfies transitivity.

As is clear from the table, $C_1(\cdot, Q)$ performs better than $C_2(\cdot, Q)$ and $C_3(\cdot, Q)$ if the fuzzy preferences are constrained; $C_2(\cdot, Q)$ and $C_3(\cdot, Q)$ perform better than $C_1(\cdot, Q)$ if the fuzzy preferences are not constrained. Through our comparative study, the rationality properties of some specific preference-based fuzzy choice functions are further understood and thus the research is of significance without doubt to the selection of choice functions in practice.

References

1. Barrett, C.R., Pattanaik, P.K., Salles, M.: On choosing rationally when preference are fuzzy. *J. Fuzzy Sets and Systems* 34, 197–212 (1990)
2. Roubens, M.: Some properties of choice functions based on valued binary relations. *J. European Journal of Operation Research* 40, 309–321 (1989)
3. Banerjee, A.: Fuzzy choice functions, revealed preference and rationality. *J. Fuzzy Sets and Systems* 70, 31–43 (1995)
4. Zhang, H.J.: A study on the fuzzification of Schwartz's theory, MA Thesis, Taiyuan University of Technology, Taiyuan (2006)
5. Guo, Y.C.: A study on rationality conditions of fuzzy choice functions, MA Thesis, Taiyuan University of Technology, Taiyuan (2005)
6. Wang, X.: A note on congruence condition of fuzzy choice functions. *J. Fuzzy Sets and Systems* 145, 355–358 (2004)

The Application of Internet of Things Technologies in Transmission Link of Smart Grid

Yu Jun and Zhang Xueying

College of Information Engineering, Taiyuan University of Technology,
Taiyuan Shanxi 030024, China

Abstract. It is the key for improving the control and management of smart grid to know how to make multiple utilization of internet of things technologies better. In this paper, the relationship between internet of things and smart grid is analyzed. This paper shows that smart grid is based on internet of things. The internet of things technologies needs of smart grid are put forward. Finally, detailed application process and the corresponding technologies are proposed aiming at the transmission link of smart grid.

Keywords: Internet of things, smart grid, transmission, perceiving information, information transmission, on-line monitoring, assistant decision.

1 Introduction

Internet of things (IOT) has been given wide attention and had the large number of applications in many fields, as a new generation of information technology. Smart grid (SG) plan first announced the project by State Grid Corporation of China. Comprehensive building of the strong SG based on UHV grid as backbone frame and coordination development of grid at all levels in May, 2009. It will become the self-directed innovation, the international leading SG which including information centered, automation, man-computer interaction. The introduction of IOT technology to SG has make good use of IOT with the actual situation of grid; it is the only road for development of SG based on IOT. [1,2]

State Grid Corporation of China specially formulated for the SG to ensure comprehensive arrangements for ASIC (Application Specific Integrated Circuit), application system development, standard system, safety of information, wireless wideband communication, software platform, testing technique, experimental technique and so on. State Grid Corporation of China strive for a breakthrough in the IOT core technologies technology in application of grid system, have influential achievements in scientific research, become influential R&D centers and industrial development zone in the next 3 years.

2 SG Based on IOT Technology

The IOT has three essential factors, by which original conception in 1999 is overall perception, reliable transmission, intelligent processing. It obtains the information of

objects through RFID, sensors, two-dimensional code and so on; the information is conveyed with expedition by telecom network and internet; analysis and processing the large quantity of information with the intelligent computing technologies (e.g. fuzzy recognition, cloud computing) for intelligent control [3].

It is can be realized that equipment condition will be predicting, preventing and regulating by the important parameters of the each link of grid on-line monitoring system. Assistant decision of transmission line and intelligent decision of power distribution depends on supervisory information. Interaction between grid Operator and consumer, developing new value-added services will be the important part of core tasks for construction of SG. The fulfillment of intelligent task should adhere to IOT with a penetrating perception, reliable information transmission, perfect network architecture, intelligent management information system and multi-stage data efficient processing.

IOT takes up the dominant place in SG because of its special advantages. It could ensures real-time, accurate and comprehensive data transmission in the important link of power generation, power transformation, power advice on Communication Standards for SG transmission, power distribution and using electricity.[4] There will be the obvious fusion trend of SG and IOT. The application of IOT technology in SG is introduced in Fig.1.

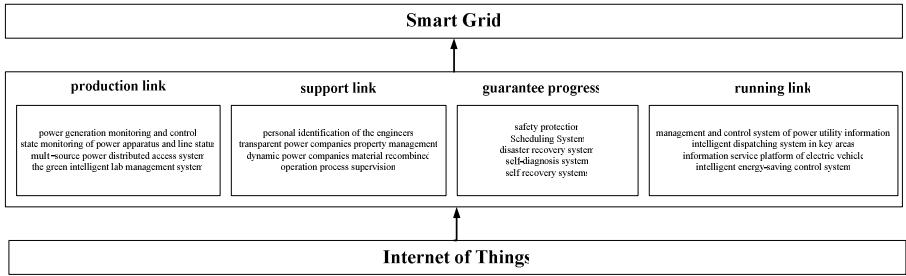


Fig. 1. Application of IOT Technologies in SG

3 IOT Based on Requirement in SG

SG based on integrated and high-speed bidirectional communications, it embraces sensing measurement technology, advanced equipment technology, advanced control strategies and group decision support systems for the running of grid with dependable, safety, economy, and high efficiency. The main character include self-healing, prompting and resist attack, power quality to meet the requirements of consumers, multi-source power distributed access, starting electric power market, efficient operation of electric power assets.

The kernel of SG lies in an intelligent internet information service system which has intelligent analyzing and adaptive various energy sources distributed access grid, and this concept from open distributed management platform.[5] This management platform could real-time collect and control of the grid data and the monitor the

residential electricity characteristics, enhancing reliability of grid and energy transfer efficiency by the best model of programme and adjustment.

SG need Technological Support of IOT in every respect (the access and testing in power generation stage, manufacturing management, safety-evaluation and supervision in power transformation stage , automation of distribution power, power monitoring, information collection of energy consumption, marketing) , the 80 percent of business of State Grid Corporation of China is closely related to IOT.

3 The Application of IOT Technologies in Transmission Link of SG[6,7]

IOT technology can improve the intelligence and reliability level of transmission segment. The collection, transmission, processing, and applying for the production data in this link of SG all need the support of IOT technology. [8,9] The application of IOT technologies in transmission link of SG has been shown in Figure 2.

Things Transport Links in the Application of Transmission Network			
Data acquisition	Data transmission	Data processing	Data application
Conductor temperature monitoring Conductor sag monitoring Equivalent thickness of ice monitoring Conductor galloping monitoring Tower tilt monitoring Insulators contamination monitoring Micro-meteorological monitoring	Home network Local area communications network On-site local area network Wide area network	Data mining Data warehouse On-line analytical processing Distributed computing Cloud computing	Digital simulation technique Visualization technology Common information platform Control decision technology Digital power equipment

Fig. 2. The Application of IOT Technologies in Transmission Link of SG

3.1 Data Acquisition

Data collection in the SG of power transmission focus mainly on several problems associated with: conductor temperature, conductor sag, equivalent thickness of ice, conductor galloping, tower tilt, insulators contamination, micro-meteorological monitoring.

(1) Conductor Temperature Monitoring

In order to prevent the transmission lines and the hardware to overheat, the SG monitors their temperature with platinum resistance or thermistor sensors. Sensors are installed in the important lines which need to increase the transmission capacity and important cross section which cross the main trunk railways, highways, bridges, rivers, seas and so on.

(2) Conductor Sag Monitoring

For leaving a safe distance between the operating transmission lines and the ground or objects under the lines, the conductor sag should be monitored by equipments (e.g. laser sensors). These sensors are installed on the lines needed to verify the new conductor sag characteristics and the track sections which may be on the blink due to for lack of a safe distance. This can be a condition monitoring system to provide early warning information.

(3) Equivalent Thickness of Ice Monitoring

In the special geographical conditions such as windward slope, pass, duct, near the large water surface, transmission lines prone to ice accretion. SG establish a mathematical model by adding load on the insulator string or putting up the real-time monitoring of line clamps, so that the equivalent ice thickness can be calculated and the rules and characteristics of the distribution of icing-coating could be master. Then related personnel could take anti-ice, melting ice and de-icing measures in time.

(4) Conductor Galloping Monitoring

Conductor galloping may damage the tower, connecting fittings and wire itself. Through the device by a number of sensors detecting conductor galloping and giving off warnings, SG can assess whether transmission lines occur injury of conductor galloping. It is in favor of mastering the rules and characteristics of the conductor galloping and putting forward preventive measures.

(5) Tower tilts Monitoring

At the bad geological section (e.g. gob, settlement areas), poles and towers of transmission grid prone to lean. Getting the two-dimensional in-formation of tower by means of dual-axis tilt sensor, the line tilt, lateral tilt and integrated tilt can be monitored. And the characteristics of the tower tilt could be grasped to avoid affecting the circuit operation.

(6) Insulators Contamination Monitoring

The insulator surface may have the accumulation of sediment gradually under the influence of atmospheric environment (e.g. sulfur dioxide, nitrogen oxides and particles of dust). After monitoring the air near the insulator by using salt density optical fiber transducer, mathematical model could be found to calculation of ESDD. So that contamination warning, line cleaning, pollution areas mapping and other work could be carried on.

(7) Micro-Meteorological Monitoring

Breaking down of tower, tripping or other accidents may be caused by terrible weather (e.g. winds, heavy rain) occurred along the transmission line. Copious sensors should be arranged in the area with complicated weather to realize real-time monitor of wind speed, wind direction, temperature, and humidity and so on. This information could accumulate weather data to provide the basis for the route planning.

3.2 Data Transmission[10]

According to the application range in the SG system, network structure for data transmission includes four levels: home network (HAN), local area communications network (LAN), on-site local area network (FAN) and wide area network (WAN).

HAN connects smart meters and smart appliances through a gateway or a user entrance to realize remote meter reading and provide optical access for home users. It allows users to participate in the electricity market according to the needs of power companies. FAN connects equipments among the field of production and control for data collection, detection and control. And the tag data of inspection equipments could also be captured. LAN connects smart meters and data concentrators. Data concentrator can collect the measurement values and information of nearby meters immediately or in accordance with the pre-set time. They can also relay commands and information from data center to downstream meter and user. At the same time, data concentrator is connected to data center through the WAN.

In data transmission network, according to the specific requirements of communication, different media could be used to implement communication, such as power line carrier communication (PLC), Zigbee, WiMax/WiMedia, WiFi, the Internet (IPv4 and IPv6), optical Ethernet, Third-generation (3G) wireless voice and data communications, wireless communications, and so on.

3.3 Data Processing

(1) Data Warehouse

With the increase in the amount of data, data must be stored with a unified form to access to information needed by decision-making more convenient, namely to form a data warehouse. Data warehouse for SG has data collection and monitoring function, state estimation function and real-time data and historical data storage function. On the one hand, data warehouse separates partial function of the original data acquisition, energy management systems, dispatcher training simulation system. On the other hand, data warehouse becomes a link between the main applications.

Centralized data warehouse in real-time data storage of grid operation, historical data, network equipment, data and other public data, reducing data redundancy. It reduces data redundancy that Real-time data and historical data of grid operation, data of network equipment and other public data are centralized in data warehouse. [11]

(2) Data Mining

Data mining dedicates to data analysis and understanding and reveals hidden knowledge of the data, which digs out the hidden information from a large number of data with noise.

It is one of the advanced means of data analysis currently. Data mining include classification model, clustering model, time series model, relational model, sequence mode. Equipment failures on the grid, daily data and operating data are analyzed by the intelligent data analysis system of grid, with a variety of algorithms and models in the fields of data mining. According to the safe operation of power grid, record data associated with analyzing factors are extracted and loaded into the data warehouse, by way of cleaning up a large number of the initial data. Then these data are dealt with

corresponding algorithm for mining to get the knowledge needed. It provides theoretical support for the insurance of power grid safe operation. [12]

(3) On-Line Analytical Processing (OLAP)

On the basis of the data warehouse, more in-depth data analysis should be made through the online analytical processing to provide flexible means of data query and report generation for Grid Company.

(4) Distributed Computing

SG is informative and these data cover a wide field. Fully utilizing idle calculating and data processing ability of computer, distributed computing can break the problem needed powerful calculation capability into many small parts. These parts are separated into geographically dispersed computers via network to be managed. Finally, the results added together to build up the analytical result.

(5) Cloud Computing

Based on a computer network, the platform of cloud computing provide unified management and dynamic allocation of a variety of computing resources. Cloud computing centers the data and uses virtualization method to integrate processing ability of large numbers of server cluster distributed in the network, and provides users with safe, reliable and convenient data services applications based on service-oriented architecture. Users can use any terminal equipment connected to the network to access these services anytime, anywhere. [13]

3.4 Data Application

(1) Digital Simulation Technique

Digital simulation technique is the calculation and software system which includes condition assessment, risk assessment, control and optimization. It can not only make a rapid response to emergencies, but also provide a powerful mathematical support to the power grid. Based on the information acquisition (via sensor and measurement), communications, and information platform, it could be realized that model all equipments (power generation, transmission and distribution, electric consumption, power conversion) and the network. Using supercomputing and the concrete structure of the power grid, digital simulation technique could be achieved by the way of distributed computing or cloud computing. Combining expert system and the simulation model, the optimum scheme under different grid conditions can be obtained. The possible problems (e.g. the deficiency of stability margin, the deficiency of damping and spare) could be predicted, and the corresponding scheme could be provided in advance. Stability, safety, reliability and the overall efficiency benefits of the power grid could be improved accordingly.

(2) Visualization Technology

Traditionally, on-site inspection relies mainly on the means of image and talkback during the process of high tension wires construction, in which real-time and accuracy are hard to guarantee. Visualization technology takes all kinds of information together to form visual information, facilitating operator for analysis and decision. Thereby, working efficiency and decision accuracy can be greatly improved. The integration

and integrated application of aforementioned foundation technologies can form various function systems in SG (e.g. new generation energy management system, new generation equipment monitoring system and market operation system) to realize the common targets such as modern power grid safe, reliable, economic, and high-quality operation.

(3) Common Information Platform

Common information platform includes grid data centers, structure of distributed data center and compatible software, and can realize data dynamic sharing, large capacity and high-speed access, standby redundancy and accurate data difficulties, etc. Based on this platform, grid chain including power generation, transmission, distribution and users could be analyzed, as well as network information including national, regional, regional, local area. And generation scheduling, planed outage, asset management, production optimization, risk management, load Management and marketing management could be formed.

Personalized visual interface is provided by using video and audio technology such as graphic display, 3d animation, speech recognition, touch screen and geographic information system (GIS).

(4)Control Decision Technology

Decision support system can identify and determine real-time problems in power grids. These problems are analyzed using repository and scientific reasoning to propose solutions to problems and decision support. Corresponding system conditions, various alternatives and their Possibility could be displayed to operating personnel. It can also be used for demand side management (DSM) and the response to user demand. [14]

(5)Digital Power Equipment

Using of sensor technology, simulation technology, communication technology and embedded systems, electrical equipment can be upgraded to be digital intelligence, which could enable operators to accurately identify the operational status, operational capabilities, overload, and the possibility of failure of electrical equipment. And life of these devices could be assessed. Digital power equipment contains transformers, chokes, switches, transmission and distribution lines can also include digital substation. Key to achieve digital power equipment is to use the appropriate sensor technology and accurate simulation models and assessment models.

4 Conclusions

It is the natural sequence to development of information and communication technology that IOT technology used in SG. It could integrate infrastructure resources in power system effectively, improving the level of power system information and the utilization ratio of establishment. As an emerging technology, IOT breaks the conventional thinking in communication field, which integrate concrete, cables, towers, chips and network. All of the concept, technology and application forms of IOT are in line with SG development needs.

Acknowledgments. Thank you for the support of the special innovation and entrepreneurship project of the university students in Taiyuan Shanxi Province 2011 (The Application of Internet of Things Technology in Transmission Electric Power Network).

References

1. State Grid Corporation of China. Smart grid technology standard system programming, Beijing (2010)
2. State Grid Corporation of China. Smart grid critical equipment (system) development programming, Beijing (2010)
3. Yang, W., Ruilan, F.: The Application of Internet of Things and Information Communications Technology. *Data Communication* (1), 12–15 (2011)
4. Ren, Y., Cao, F., Tang, X.: Advice on Communication Standards for Smart Grid. *Automation of Electric Power Systems* 61(3), 1–4 (2011)
5. Cheng, S., Li, X., Zhang, Z.: Entire-grid-area Information-sharing and Integrated Applications in United Information System for Smart Grid. *Proceedings of the CSEE* (1), 8–14 (2011)
6. Liu, Z.: Smart Grid Technology. China Electric Power Press, Beijing (2010)
7. Xiao, L.: Smart Grid Technology System. *Electrical Engineering* (3), 1–3 (2010)
8. Chen, X., Li, X., Wang, H.: Applications of Internet of Things in Smart Grid. *North China Power* (3), 50–53 (2010)
9. Li, X., Liu, J.: Internet of Things Technology and Its Applications for Smart Grid. *Telecommunications Network Technology* (8), 41–45 (2010)
10. Xu, L.: Communication Network Framework and Key Technologies for Smart Grid. *Electrical Engineering* (8), 16–20 (2010)
11. Wang, W., Zeng, W., Dai, W.: An Integrated Automation System for Electric Power Dispatching Based on Data Storage. *Automation of Electric Power Systems* 12, 67–70 (2003)
12. Liu, F.: Research and Implementation of Intelligent Analysis for Electric Grid Data. Northwest University, Xi'an (2008)
13. Dispatching. *East China Electric Power* 38(6), 800–803 (2010)
14. Li, N., Ni, Y., Sun, S.: Survey on Smart Grid and Relevant Key Technologies. *Southern Power System Technology* 4(3), 1–7 (2010)

Assess of the Flicker Caused by Electric Arc Furnace Using Simulation Method^{*}

Wang Yufei¹, Hua Xue¹, and Yu Xiao²

¹Department of Electrical Engineering, Shanghai University of Electric Power
200090, Shanghai, China

²Electrical equipment installation branch, Shanghai Boye Group Corp. LTD.
200090, Shanghai, China

w.powerquality@gmail.com, xiaobaoye@126.com

Abstract. In order to assess the flicker caused by electric arc furnace, the simulation method to assess flicker and its application in electric arc furnace power system is studied. Firstly, based on the Flicker Meter functional and design specifications recommended by IEC, an IEC Flicker Meter is designed concretely; Secondly, an IEC Flicker Meter model is established in SIMULINK, and a special M file which can be used to calculate the short term flicker sensation indicator (P_{st}) is compiled, according to the IEC criterion, the Flicker Meter model is tested; Finally, assess of the flicker caused by an actual electric arc furnace is achieved, the instantaneous flicker sensation $S(t)$ in the point of common coupling is gained. The results show that application of simulation method to assess the flicker caused by electric arc furnace is an effective way.

Keywords: Electric arc furnace, voltage fluctuation and flicker, Flicker Meter, simulation.

1 Introduction

When impact load of high power connected to power grid of relatively small capacity, such as electric arc furnace, mine hoist, heavy merchant mill, heavy electric welding machine as well as electric locomotive, voltage fluctuation will be caused, which will undermine other electric equipments which are connected to the point of common connection. With the wide application of the above mentioned impact load of high power, as ref. [1], voltage fluctuation and flicker has become one of the key indicators to measure the power quality of the grid.

Electric arc furnace (EAF) is not only a major facility in steel factory but also one of the super impact loads of single capacity in the power supply system. The components of voltage fluctuation frequency due to EAF is mainly distributing in 4~14Hz band,

^{*} Project supported by National Natural Science Foundation of China(50977055,51007053), Leading Academic Discipline Project of Shanghai Municipal Education Commission (J51301), Shanghai Key Scientific and Technical Project(09160501600, 10dz1203100, 10110502100)and Shanghai University Scientific Selection and Cultivation for Outstanding Young Teachers in Special Fund(sdl09005).

which is precisely in the visually sensitive areas of human beings. Therefore, among all sorts of impact load, EAF has the most severe impact on voltage fluctuation and flicker as shown in ref. [2]. The regulations from home or abroad on voltage fluctuation and flicker limits are specific about EAF. Generally, the terms can be applied to all the other types of loads if they can meet the requirements of EAF. In order to improve or inhibit the voltage fluctuation and flicker caused by EAF, some relevant compensation facility such as SVC which is needed to be used in the power grid. Because accurate flicker values are needed for the research and setting of these devices, the accurate measurement or estimation to flicker value is the prerequisite for solving the problem of voltage fluctuation and flicker.

The flicker severity caused by EAF can be directly measured by Flicker Meter or estimated by simulation method. As shown in ref. [3-7], the research and application of Flicker Meter has been the focus of experts from home and abroad. In recent years, there are literatures (ref. [8-11]) on flicker simulation method emerged, but most of them are only focusing on the simple simulation method research. And literatures of further studies on application have been rarely involved, so they are not supposed to guide the practical application. In addition, the accurate signal of voltage fluctuations which can be obtained through simulation system is a prerequisite for the application of simulation method. The accuracy of the signal depends on the mathematical model of load characteristics. The signal can also be obtained through analysis of field measurement data. When the waveform of voltage fluctuations is recorded, waveform of any time-domain can be divided into a series of residential section of signals which are input to the flicker simulation measurement system for data processing. And finally get the results of instantaneous flicker sensation level $S(t)$ and other short-term flicker value P_{st} .

2 IEC Flicker Meter

The designing standards of the Flicker Meter recommended by IEC only provides schematic diagram and design specifications, without mention of the specific design details and design parameters. Therefore, understanding the physical sense of the input and output signals of each part in the flicker measurement system correctly and setting specific technical details in each part become the priority of the research of flicker measurement system. The schematic diagram of the Flicker Meter recommended by IEC61000-4-15 (ref. [12]) is shown in Figure 1.

In figure 1, the frame 1 is an input adapter link with fluctuant voltage signal input. It is used to reduce the input voltage of different levels to the level applying to internal circuits.

The frame 2-4 simulate of the reaction of light- eye-brain to the voltage fluctuation. Among which, frame 2 simulate the function of light, realizing square demodulation to the voltage fluctuation component and obtaining the voltage signal in linear relation with component of the voltage fluctuation.

The band-pass weighting filter part of the frame 3 is made up with a band pass filter and a visual sensitivity weighting filter in series, which reflect the sensitivity of the illuminance change of tungsten filament lamp of 60w, 230v in the voltage fluctuations

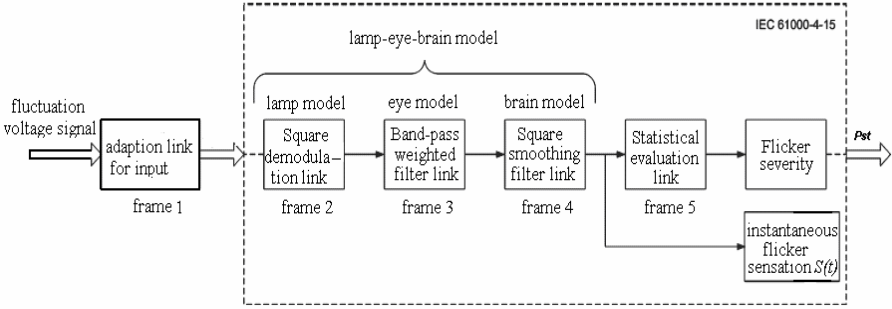


Fig. 1. Schematic diagram of IEC flickermeter

of different frequencies. The band-pass filter consists of a high-pass filter and a low-pass filter with the frequencies 35Hz and 0.05Hz correspondingly. The transfer functions of the two filters are listed in equations (1) and (2). The transfer function of the first-order high-pass filter with the cutoff frequency of 0.05Hz is as follows.

$$H_p(s) = \frac{s/\omega}{1 + s/\omega} \tag{1}$$

Where, $\omega=2\pi 0.05$. The transfer function of the low-pass filter with a cut-off frequency of 35Hz which select the 6-order Butterworth filter is shown in equation (2).

$$Bw(s) = [1 + \sum_{i=1}^6 b_i (\frac{s}{\omega})^i]^{-1} \tag{2}$$

Where, $\omega=2\pi \times 35$, $b_1=3.864$, $b_2=7.464$, $b_3=9.141$, $b_4=7.464$, $b_5=3.864$, $b_6=1$.

The visibility weighting filter in frame 3 is used to simulate the frequency selectivity of the human eyes. The center frequency is 8.8 HZ and the transfer function is shown in equation (3).

$$F(s) = \frac{k\omega_1 s(1 + s/\omega_2)}{(s^2 + 2\lambda s + \omega_1^2)(1 + s/\omega_3)(1 + s/\omega_4)} \tag{3}$$

Where, $k=1.74082$; $\lambda=2\pi 4.05981$; $\omega_1=2\pi 9.15494$; $\omega_2=2\pi 2.27979$; $\omega_3=2\pi 1.22535$; $\omega_4=2\pi 21.9$.

Frame 4 contains a square and a low-pass smoothing filter with the time constant of 300ms. The transient and nonlinear response and memory effect to the illuminance of the light are simulated for the light-eye-brain parts. The low-pass smoothing filter transfer function is shown in equation (4).

$$L_p(s) = \frac{1}{1 + 0.3s} \tag{4}$$

It should be noted that in order to obtain the instantaneous visual sensitivity $S(t)$, which reflects the flicker of the voltage fluctuations of human's, further processing for the output signal from the first order low-pass smoothing filter is needed.

Cumulative Probability Function (CPF) method is used in frame 5 to make statistical estimation of the flicker. Flicker value of P_{st} in short time (10 min measurement time) is commonly used as indicator to flicker evaluation, the value of which defined by IEC is shown in equation (5).

$$P_{st} = \sqrt{0.0314P_{0.1} + 0.0525P_1 + 0.0657P_3 + 0.28P_{10} + 0.08P_{50}} \quad (5)$$

Where, $P_{0.1}$, P_1 , P_3 , P_{10} , P_{50} are all visual sensitivity values $S(t)$ obtained within 10 minutes by using CPF values exceeding 0.1%, 1%, 3%, 10% and 50% of the time respectively.

3 Simulation Models and Verification of the IEC Flicker Meter

According to the transfer function and specific parameters in each part of the above IEC Flicker Meter, the Flicker Meter simulation model is established using SIMULINK, which is showed in Figure 2. The system input of the figure is the component of the amplitude modulated wave signal of the fluctuation voltage, the output of which is the instantaneous flickering sensitivity $S(t)$.

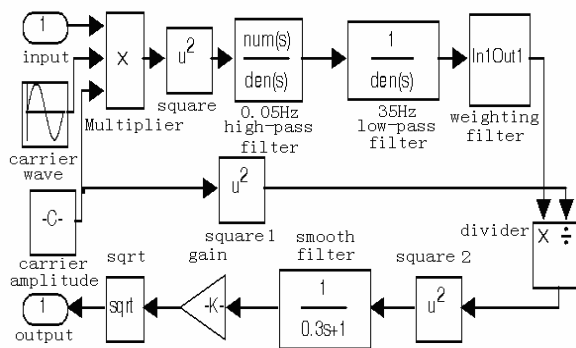


Fig. 2. Simulated model of Flicker Meter

To verify the validity of simulated model of the Flicker Meter, the input signal of the voltage signal is assumed as $u(t)$, which is shown in (6).

$$u(t) = 230\sqrt{2}[1 + m \cos(\Omega t)]\cos(\omega t) \quad (6)$$

Where, m is the ratio between the peak amplitude modulated wave value and the peak rated voltage value, whose value equals to half the value of the voltage fluctuations $\Delta U/U$; Ω is the frequency of the amplitude modulation wave angular whose value equals to $2\pi k$, where k is the wave frequency and ω is the frequency of the power frequency angular.

With the system simulation testing time of 10 minutes, and the data of the voltage fluctuation of the sine amplitude modulates wave when $S(t) = 1$ according to IEC, the input signal of the voltage fluctuation is constructed to verify and simulate the value

of $S(t)$ shown in table1. According to the calculated relative errors between the simulation results and the theoretical value of $S(t)$, the relative errors remains between $\pm 5\%$ that is well compliant with the accuracy requirements listed in the standard of IEC. Therefore, the validity and availability of the simulation system model of the Flicker Meter is proved.

Table 1. Proof of the IEC Flicker Meter model

Fluctuation frequency	Fluctuation Value (%)	simulation value $S(t)$	Relative error (%)
1.0	1.432	0.952	-4.8
3.0	0.654	0.985	-1.5
5.0	0.398	0.997	-0.3
6.0	0.328	0.998	-0.2
6.5	0.300	0.998	-0.2
7.0	0.280	0.998	-0.2
7.5	0.266	0.998	-0.2
8.0	0.256	0.998	-0.2
8.8	0.250	0.998	-0.2
9.5	0.254	0.998	-0.2
10.0	0.262	0.997	-0.3
10.5	0.270	0.997	-0.3
11.0	0.282	0.996	-0.4
11.5	0.296	0.995	-0.5
12.0	0.312	0.997	-0.3
13.0	0.348	0.997	-0.3
16.0	0.480	0.997	-0.3
18.0	0.584	0.999	-0.1
20.0	0.700	0.999	-0.1
25.0	1.042	1.011	1.1

By outputting the sample value of $S(t)$ obtained from the flicker testing system to “sw.mat” file by the “To File” module, the P_{st} value is calculated based on the method of CPF and M file compiled by MATLAB.

4 Evaluation of the Flicker Severity Caused by Electric Arc Furnace Using Simulation

By applying the established model of Flicker Meter and the stochastic model of the arc furnace proposed by the author, the voltage fluctuation and the flicker severity caused by the arc furnace can be evaluated in a simulate way. Then the instantaneous flicker sensitivity value $S(t)$ of a common connection point (PCC) of the power supply system with arc furnace in a certain steel mill is obtained.

Based on the sampling value $S(t)$ from the “sw.mat” file, five eigenvalues of the instantaneous flicker sensitivity are obtained when running the M file. The five P_{st} are as followed: P0.1, P0.2, P0.3, P0.4, P0.5, are respectively 1.7130, 1.6299, 1.5301, 1.3305, 1.0146. The value of P_{st} obtained by composing above figures into equation (5) is 0.8328. It is obvious that the flickering value of P_{st} in short-time caused by the arc furnace is a little higher than international standard (On the condition of HV, the limited value of P_{st} is 0.8). Therefore, proper measures are needed to take to inhibit the question of voltage fluctuation and flicker caused by the arc furnace.

5 Conclusion

Based on the flicker measurement principle recommended by IEC, the transfer function and specific parameters of each session in the IEC Flicker Meter are designed to facilitate understanding the physical meaning of the input and output signals of each session in the IEC Flicker Meter. In the flicker measurement system model established in SIMULINK, the simulated value of $S(t)$ is obtained by constructing the input voltage fluctuation signal according to the inspection standards recommended by IEC. And the relative error between the simulated value and the theoretical value keeps below $\pm 5\%$, which is absolutely accords with the requirement mentioned in the standard of IEC. Meanwhile, the experimental results show that this simulated system model of the Flicker Meter is valid and effective. With reference to the stochastic models of the AC arc furnace, flickering severity caused by arc furnace is researched. It is shown that it's available and effective to evaluate the flicker severity caused by arc furnace by using the simulation method, which also provides a new effective way to evaluate the flickering severity.

References

1. Haixue, L.: Introduction of the new national standard of Power quality voltage fluctuation and flicker. For Electricity 18(6), 4–7 (2001)
2. Sun, S.: Voltage fluctuation and flickering. China Electric Power Press, Beijing (1998)
3. Cristina, G., Thomas, C., Van den, K.J., et al.: Development of a Flicker Meter for grid-connected wind turbines using a DSP-Based prototyping system. IEEE Transactions on Instrumentation and Measurement 55(2), 550–556 (2006)
4. Gutierrez, J.J., Ruiz, J., Ruiz de Gauna, S.: Linearity of the IEC Flicker Meter regarding amplitude variations of rectangular fluctuations. IEEE Transactions on Power Delivery 22(1), 729–731 (2007)

5. He, J., Huang, Z., Li, Q.: Development of Flicker Meter and application in electrified railways. *Journal of Southwest Jiao Tong University* 39(2), 217–221 (2004)
6. Jia, X., Zhao, C., Xu, G., et al.: Error analyzing of IEC Flicker Meter and its improving design. *Transactions of China Electrotechnical Society* 21(11), 121–126 (2006)
7. Bucci, G., Fiorucci, E., Landi, C.: A digital instrument for light flicker effect evaluation. *IEEE Transactions on Instrumentation and Measurement* 57(1), 76–84 (2008)
8. Bertola, A., Lazaroiu, G.C., Roscia, M., et al.: A Matlab-Simulink flicker model for power quality studies. In: 11th International Conference on Harmonics and Quality of Power, pp. 734–738 (2004)
9. Qi, B., Xiao, X.: Simulation of IEC Flicker Meter. *Journal of North China Electric Power University* 27(1), 19–24 (2000)
10. Mazadi, M., Hosseinian, S.H.: Flicker Meter simulation to use in power system analysis software. In: 39th International Universities Power Engineering Conference (UPEC), pp. 917–923 (2004)
11. Ma, Y., Liu, L., Zhang, J., et al.: Research of digital Flicker Meter based on IEC standard. *Proceedings of the CSEE* 21(11), 92–95 (2001)
12. IEC 61000-4-15. Flicker Meter-Functional and design specifications

Research to Prevent the Anti-charging Incident of the Voltage Transformer in the Substation

Hui-Ying Chen, Zhi-Peng Wang, and Mu-Qin Tian

College of Electrical and Power Engineering, Taiyuan University of Technology,
West Yingze Street No.79, 030024 Taiyuan, Shanxi, China
tianmuqin@tyut.edu.cn

Abstract. The voltage transformer of the power system is a device which converts the high voltage of the primary side into the low voltage required by secondary device, and the primary side connects to the primary system, the secondary side connects to the measuring instrument, relay protection equipment and so on. This paper Introduces a kind of air switch with the ability of preventing the anti-charging occurring in the voltage transformer from secondary side of the to primary side in the ordinary air switch used in the practice. It will automatically handle the anti-charging phenomenon and protect the safety of the staff and equipments more effectively.

Keywords: Voltage transformer, anti-charging, tripper, automatic processing.

1 Introduction

There are many reasons to cause the anti-charging from secondary side to the primary side of the voltage transformer. They are as follows: the parallel switching when the multi-buses running, wrong connecting the power bus into the blackout bus of the secondary side when maintenance personnel works in the secondary side, connecting the power bus and the blackout bus due to the secondary parasitic circuit caused by the disordered wire-connect of voltage transformer secondary side. In order to fundamentally solve the incident, this paper presents a kind of air switch with the ability for preventing the anti-charging.

2 Accident Analysis

It is discovered that the accident of the voltage transformer anti-charging mainly occurs in the parallel switch operation in the practical work, the following will analyze the anti-charging accident base on the operation.

The parallel switch operation of the voltage transformer secondary side means that voltage transformer secondary side can be switched parallel in order to ensure that the primary side devices work continually and secondary side devices such as measuring

instrument, relay protection equipment and so on get voltages corresponding to the primary voltage when one of the voltage transformer working in double-bus or sectionalized single-bus configuration will be maintained for any reason.

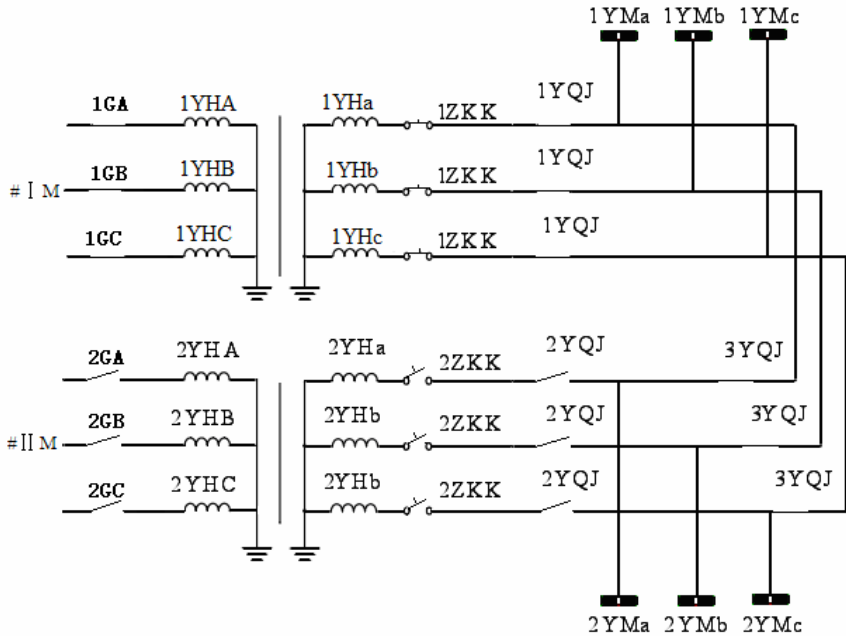


Fig. 1. The parallel and switch operation of the voltage transformer three-phase AC circuit

The Fig.1 shows the parallel switch operation of the voltage transformer in the three-phase AC circuit where YM is small bus, YH is voltage transformer, ZKK is air switch, YQJ is voltage switching contact, G is linked switch. It is supposed that 2YH will be Maintained outage, normally the air switch of 2ZKK is disconnected in the first, then 2YQJ is switched off through the linked switch of 2G, the contact 3YQJ is closed at last. In this way, the small busbar of 2YH side don't lose power and the circuit attains the effect of the parallel switch operation.

One of functions of the air switch is to prevent the accident of the anti-charging, but the function must be achieved by the manual operation, so it is easy to cause the false operation or not operation in the practice. If the linked switch doesn't make the YQJ contact completely disconnect, the accidents will happen. By the fig.2 it is known that the voltage of the secondary side inversely transfers to the primary side of the II bus's voltage transformer, it will lead to a capacitive current on the primary side equipment of the II busbar. A high-current may be produced by this current in the secondary circuit once again. The air switch only has the function of the overcurrent protection and undervoltage protection, and this moment the 1ZKK and 2ZKK of the air switch will all trip because of the high-current. The situation will cause the serious harm for the overhaul workers on the primary side and the equipment of the

secondary side. At the same time, If the high-current passes the zero sequence voltage circuit long time, it will burn out the zero sequence voltage parallel relay in the voltage transformer's parallel plug-in because there is not the air switch in the zero sequence voltage circuit[1-2].

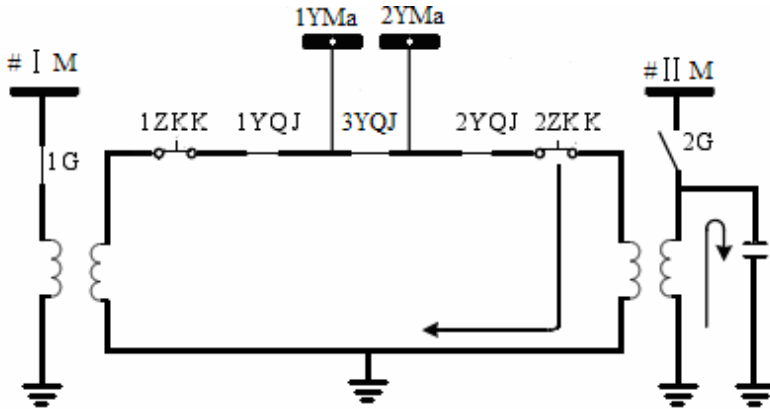


Fig. 2. The schematic of the harm of the voltage transformer anti-charging

3 The Present Preventive Measure

At present, the phenomenon of the anti-charging is usually prevented by the following measure: switching off the air switch of the secondary side of the voltage transformer, taking off the fuse of the secondary side of the voltage transformer and setting up the ground line on the conductor of the primary side of the voltage transformer and so on.

However, there are mistakes in implementation of the security measures in the practice. The common failure of the ground line such as the installation error, the missing installation, the non-standard installation led to the poor contact and so on. The common failure of the switching operation such as the secondary air switch being not opened, the insurance being not removed, the air switch having been pulled off but actually it being not been disconnected, the secondary side auxiliary contact having bad conversion.

4 A Kind of Air Switch with the Ability for Preventing the Anti-charging

Through the above analysis with the technique actual and the engineering requirement, the solution of the preventing the anti-charging accident has been proposed.

1) The wiring should be connected strictly according to the relational procedure in the design of the secondary electrical, and artificially short or fumble connecting of some contacts of the secondary equipment are not allowed.

2) The corresponding air switch of the secondary side of voltage transformer must be cut when cutting off the voltage transformer.

This paper presents a kind of air switch with the ability for preventing the anti-charging, which means that we set a kind of tripping device with a function of anti-charging. The Security risks have been fundamentally solved to achieve the purpose of the automatic processing of the accident.

The air switch consists of three parts: the contact and the system of the arc-control (the part of opening and closing circuit), all kinds of tripper (detecting the abnormal state of the circuit and then making response, that is the part of the protective action), the mechanism of the operation and the automatic tripper (the part of the middle contact). The Fig.3 is schematic diagram of the air switch with the ability for the preventing the anti-charging.

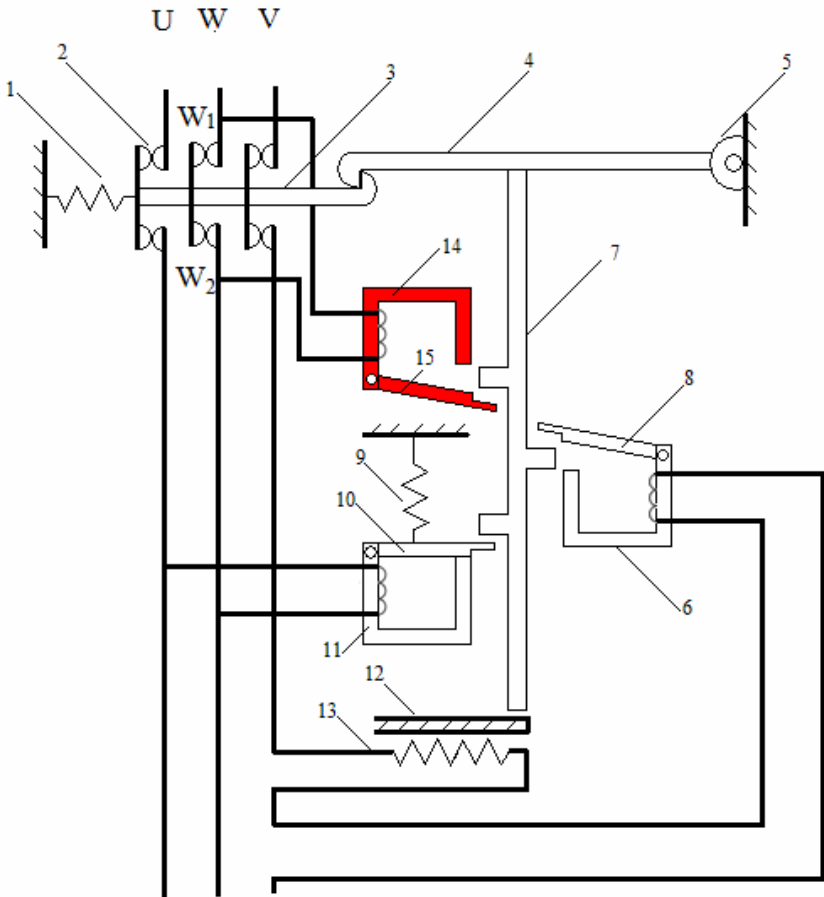


Fig. 3. The schematic diagram of the air switch with the ability for preventing the anti-charging

In the Fig.3, the contact 2 has three pairs of switches that are in series in the protected circuit. The button is pulled to the position of the “close” by the hand, and then the contact 2 keeps close condition by the locking key 3, the locking key 3 is supported by the agraffe 4. If we want to break the switch, we need pull the button to the position of “open”, and the agraffe 4 is hit by the lever 7(the agraffe 4 can rotate around the axis 5), the contact 2 is undrawn off by the spring 1, then the circuit is disconnected.

The automatic disconnection of the air switch is made up of the magnetic tripper 6, undervoltage tripper 11, thermal overload tripper 12 and the new anti-charging tripper 14.

The air switch can equip the different tripper by the different purpose.

The coil of the magnetic tripper 6 is in series with the circuit, the armature 8 is controlled by the current in the circuit, and then the air switch is disconnected through the lever 7 of being hit.

The coil of the undervoltage tripper 11 is parallel with the circuit. The armature 10 is controlled by the voltage in the circuit. The bimetallic strip is bent because the overcurrent pass the heating element 13 of the thermal overload tripper 12 when the over loading in the circuit, which is used for the protection of the over loading.

The coil of the anti-charging tripper 14 is paralleled on both ends W1W2(or other auxiliary contact position) of the air switch. The air switch is closed at both ends when the circuit properly works so it doesn't have voltage and current, the coil of the anti-charging tripper 14 doesn't act. The coil of the anti-charging tripper 14 will act when the air switch's top W1 doesn't have power but its bottom W2 has power, and the armature 15 bumps the lever 7 and hit agraffe 4, and then the circuit is disconnected by the contact 2. It can prevent the accident of the anti-charging.

At present, there are four kinds of the anti-charging tripper, it is respectively AC/DC12V~24V, AC/DC24V~48V, AC110V~127V, AC230V~400V. We found that the first two kinds have high rates of the accident of the anti-charging, the other have low rate. We also found that the smaller voltage (e.g. 1V, 6V) has high rate besides the above voltage standard, and similar dangers exist in the accident. According to the actual process and measurement accuracy, these low voltages aren't achieved by the tripper.

To the accident of the low voltage, the low tension transformer is used in the circuit. The Fig. 4 shows the low tension transformer 16 is in series with both ends W1W2. The low voltage side connects the main circuit, the high voltage side connects the tripper of the anti-charging. The existing low voltage is changed to the voltage level of the existing by the low tension transformer when there are the accident of the low voltage of the anti-charging, then the air switch is disconnected by the tripper of the anti-charging.

Finally, we can increase 5 to 6 trippers (the operation voltage is 1 V, 6 V, 12 V, 24 V, 48V and so on) according to actual situation to expand the adaptive. Not only it provides convenience for the user but also it improves the value in use.

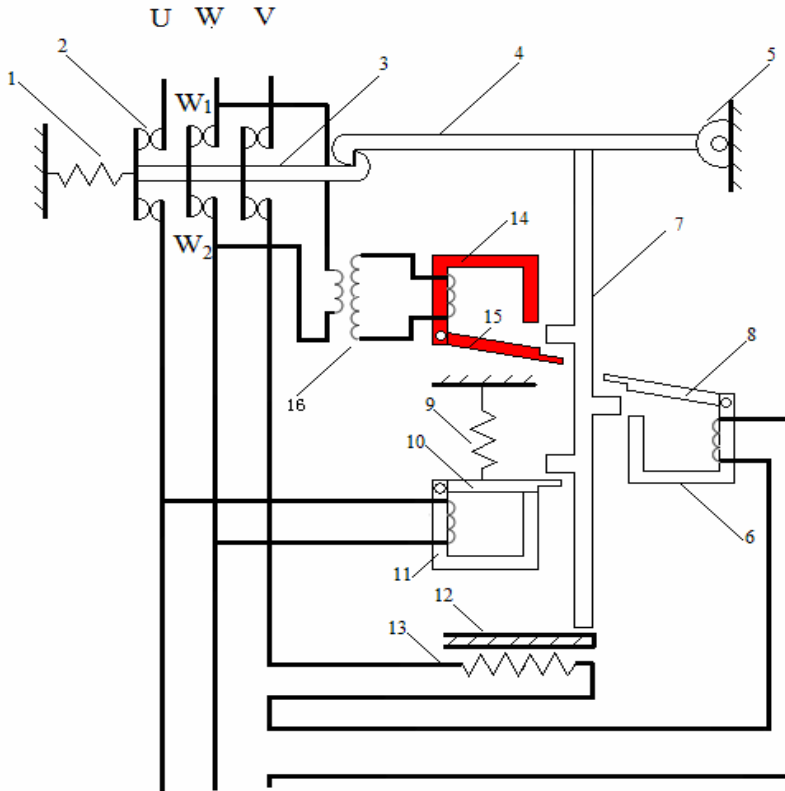


Fig. 4. The schematic diagram of the air switch with the ability for preventing the anti-charging and low tension transformer transformer

5 Conclusion

The air switch with the ability for preventing the anti-charging can improve the safety for the maintenance of outage, and advance the level of management for the power system. The accident is basically avoided by this device. It has many advantages as simple structure, convenient installation, easy maintains. This device doesn't have the standby circuit, it provides the safe assurance to the other equipments, and it also increases the value of use and promotion.

References

1. Wu, J.-h., Li, M.-c., Sun, J.-p.: Thinking of two anti-charging incidents of voltage transformer. *Power System Protection and Control* 38(13), 124–129 (2010)
2. Zhang, J.-h.: Analysis and Improvement Measures of the Secondary Reversed Feeding. *Jiangsu Electrical Engineering* 30(1), 39–41 (2011)

Research on Online Monitoring Method for Longitudinal Rip of Steel-Core Belt

Qiao Tie-Zhu, Wang Fu-Qiang, and Lu Xiao-Yu

Measuring and Controlling Technology Institute, Taiyuan University of Technology,
Shanxi Taiyuan 030024, China
qtz2007@126.com,
wayf2001@163.com

Abstract. Steel-core belt is the large transport device in coal mine and other departments, the longitudinal rip of steel-core belt will bring great losses to production and safety, this article proposed one kind of online monitoring method based on distance measuring sensor and 8051 MCU. This method is to use distance measuring sensor to real-time monitor the width of the steel-core belt, and through compiling 8051 MCU process, realize online monitoring for the longitudinal rip of steel-core belt.

Keywords: Longitudinal rip, distance measuring sensor, MCU, data acquisition.

1 Introduction

Steel-core belt conveyor is a large transportation equipment, which is widely used in mine, metallurgy, port, electric power and other department, it has some features such as high efficiency, great carrying ability and easy to realize automatic control, made it be widely used in coal production automation. However, steel-core belt in its width direction tensile strength is very low, makes the conveyor belt frequent happen longitudinal rip accidents. Steel-core belt occurs longitudinal rip for many reasons, and its main aspects including: Some rod-like material insert conveyer belt; large elongated rock fall to steel-core belt; some fixed parts of the rack to caught conveyor belt; various iron wire hitched conveyor belt and so on. Steel-core belt is expensive, it will cause significant directly or indirectly economic losses once produce longitudinal rip[1-3].

Opaque material on the steel-core belt leaking through the gap when the steel-cord belt occurs longitudinal rip. By infrared photoelectric sensor placed below the conveyor belt detects whether there are any opaque material leave out, which can detect if the tear fault occurred[4]. This detection method is structure simple and easy to install, but in some cases, after the belt is torn, the two side of the conveyor belt overlap, the cracks are sealed, result in the opaque material can not leave out, so this method will not detect the tear fault. To this detection method's deficiencies and limitations, this paper have been modified and improved its, presents a detection method based on distance measurement sensor and 8051 MCU, and design a new conveyor belt longitudinal tear online monitoring system use this method.

2 The Working Process of the Monitoring System

The width of the steel-cord belt is a constant during the conveyor belt operation , when the longitudinal tear occurs, the bandwidth will change. Through online real-time monitor the width of the conveyor belt to diagnose whether there have been occurred tear fault.

The working process of the monitoring system is: Ranging sensors, arranged on both sides of the conveyor belt, online real-time measure the width of the belt and send out the voltage signal, during the conveyor belt running. The voltage signal is filtered and amplified treatment, and through the multi-channel analog switch sent to the A/D converter, after A/D converter finishes the signal conversion, puts the digital signal sent to the 8051 MCU, after single chip calculating and processing, the data through communication interface is sent to the computer for further data analysis, image reconstruction, storage and print, and real-time shows the width of the belt in the LED screen. When the changes of the width of the conveyor beyond the set alarm threshold , which can be obtained through several tests, 8051 MCU will send the alarm to stop the conveyor running, thus prevent the expansion of the degree belt torn, and warn the technical workers overhaul. This can be achieved online real-time monitoring for the tear fault of steel-core belt.

3 Monitoring System Design

3.1 Page Numbering and Running Heads

The hardware of this system is mainly composed by the range-finding sensors parts, data acquisition circuit components, 8051 microcontroller and peripheral cricuit section, communications interface and the PC. Structure of the system is shown in fig 1.

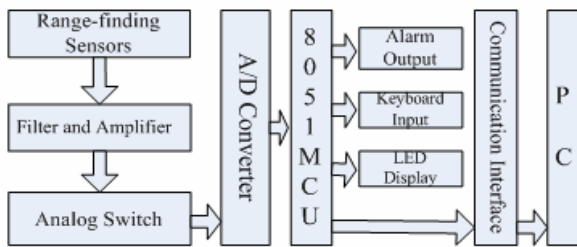


Fig. 1. Schematic diagram of the system

3.2 Page Numbering and Running Heads

Two range-finding sensors installed on both sides of the conveyor belt, two distance sensors respective detect the distance between the sensors to the conveyor belt for L_1

and L_2 , the distance between two sensors is fixed at L , then the width of the conveyor belt D can be calculated:

$$D=L-(L_1+L_2)$$

Sensors' arrangement is shown in fig 2.

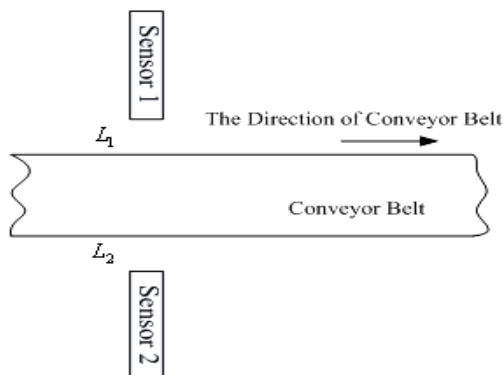


Fig. 2. Range-finding sensors installation position

3.3 The Data Acquisition Circuit

Data acquisition circuit includes the following four parts:

- (1) Filter circuit: Using a hardware filter to remove the interference part of the signal.
- (2) Amplifier: Using the operational amplifier enlarge the usefulness weak input signal to a considerable degree to match the back of the A/D converter' scale, in order to obtain the highest possible resolution.
- (3) Multi-channel analog switch: Select the several input signal to achieve time-share A/D converter.
- (4) A/D converter: Put the analog signals that come from the analog input channels in the front convert to data signals, and sent to the microcontroller. In this system, analog-digital conversion completed by the AD574A. AD574A is the 12-bit successive approximation A/D converter chip that was produced by Analog Devices, it is equipped with three-state output buffer circuit, can directly with various 8-bit or 16-bit CPU connections, do not additional logic interface circuit, and contains high-precision reference voltage and clock circuitry, it does not require any external circuit and the clock signal of to complete analog-digital conversion, the conversion time is only 35 μ s[5].

3.4 Microcontroller and Its Peripheral Circuits

This system uses Intel's high performance 8-bit microcontroller which fabricated in CMOS process, 8051 can compatible with TTL and COMS logic levels, with five interrupt sources, programmable full duplex serial port, with two layers of priority

level interrupt structure, can realize rapid pulse programming, with 4KB internal ROM and 128 bytes of internal RAM, and 8051 MCU have the advantages include: powerful control ability, simple programming, stable performance, small size, fast speed, real-time performance, low cost, take up less space, flexible installation, easy integration ect. It can make up the shortcomings that large measurement errors and poor reproducibility of previous testing equipment.

Signal through the A/D converter sent to 51 singlechip, after the microcontroller operation, sent the results into the liquid crystal to real-time show the current width of the conveyor belt, in the meantime sent the large amounts of processed data into the host computer for further processing. Because of the serial port of 8051 is a standard TTL level interface, while the serial of host computer is the RS-232C standards, in order to complete the data transmission from MCU to the host computer, need to use the MAX232 chip to achieve the communication between 8051 and PC[6]. Interface circuit is shown in fig 3. After the data through the I/O interface into the computer, then use the computer for the data to do the further processing, image reconstruction, dispaly, storage and print.

In addition, in the design process of the monitoring system in order to improve anti-jamming capability, circuit adopted the photoelectric isolation technology, space shielding technology, ect. The alarm threshold for the change of the bandwidth can obtain by test the same type conveyor belt mang times, and stored the alarm threshold in 8051, and through prepare microcontroller program, compare the current change of bandwidth with the size of alarm threshold, when the variation of the conveyor reaches the alarm threshold level, microcontroller will send alarm. In the operation process of the conveyor belt, in order to prevent the error alarm by wear and tear, should in time to change the size of the alarm threshold in the event of wear and tear.

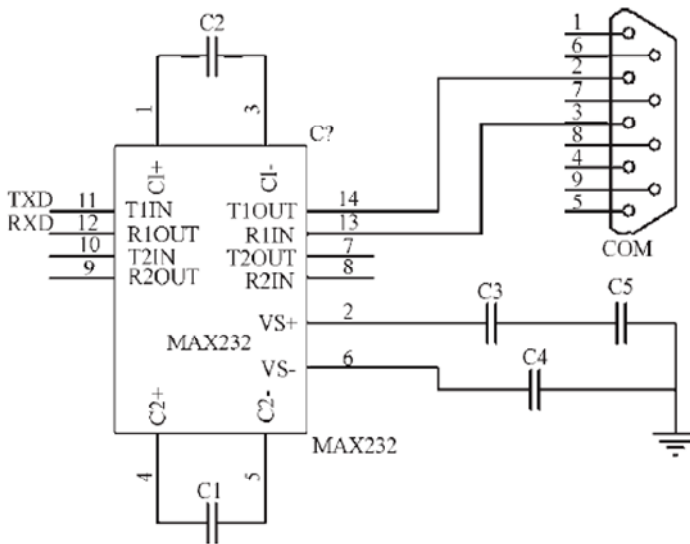


Fig. 3. The interface circuit of 8051 and host computer

3.5 The Data Acquisition Circuit

System software mainly consists initialization procedures, data collection procedures, the logical judgment main program, data shows subroutines, alarm output procedures, serial communication subroutines, ect. Fig 4 shows the block diagram of system software.

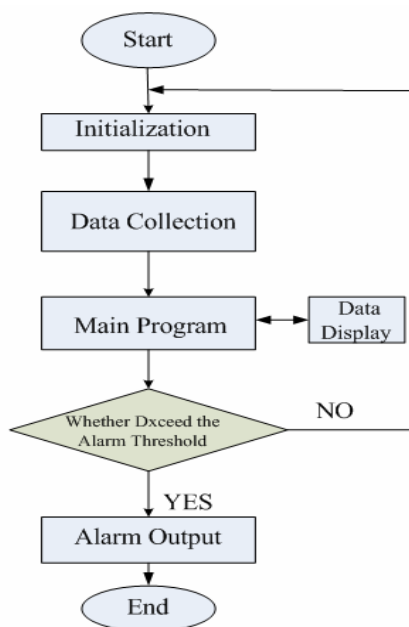


Fig. 4. Program structure diagram

4 Conclusion

This paper describes a kind of steel-core belt longitudinal tear fault diagnosis method based on range-finding sensors and the 8051 microcontroller combine, it can realize digital online monitoring, real-time display the measurment data, more limit alarm, the host computer complete data storage and print.

This detection method and the leak material detection method combined will have significant effect, it has a certain degree of research and reference for conveyor belt longitudinal rip detection technology.

References

1. Liu, Y.: Testing and monitoring longitudinal tear in conveying belt. Shan Xi Mining Institute Learned Journal 13(2), 150–153 (1995)
2. Huang, M., Wei, R.-z.: Real time monitoring techniques and fault diagnosis of mining steel cord belt conveyors. Journal of China Coal Society 30(2), 245–250 (2005)

3. Zhang, A.-n., Sun, Y.-k., Yin, Z.-h.: Research status and tendency of longitude tearing protection for belt conveyor. *Coal Science and Technology* 35(12), 77–79 (2007)
4. Qiao, T.-z., Zhao, Y.-h., Ma, J.-c.: Design on New Online Monitoring and Measuring System of Longitudinal Tearing for Conveyor Belt. *Coal Science and Technology* 38(2), 55–57 (2010)
5. Chen, Y.-k.: General data acquisition and processing system based on 80C51 microcontroller. *Machinery* 35(4), 49–51 (2008)
6. Liu, H.: Design of Data Acquisition and Communication Interface Based on 8051 Single-chip Microprocessor. *Software Guide* 7(12), 72–74 (2008)

Establishment and Analysis of Information Processing Models on Intelligent Overtaking Behavior

Xin Gao¹, Gang Dong², and Li Gao¹

¹ School of Mechanical and Vehicular Engineering, Beijing Institute of Technology,
100081 Beijing, China

² Traffic Bureau in Xiangfang District of Harbin,
150030 Harbin, China

{gaoxinbit, ligaobit}@bit.edu.cn, donggang20110201@163.com

Abstract. In order to avoid the traffic accidents caused by drivers, unmanned ground vehicles are expected to run on the road, so correct and safe driving behaviors are most important. This paper aims to give a new way of perception based on information processing for control system of intelligent vehicles, and focuses on the present overtaking ability of intelligent vehicles, meanwhile implies the ability of recognizing traffic environment. First we built intelligent driving behavior models with construction perception and pattern recognition on prototype matching; second, based on "Future Challenge" contest we investigated the features of velocity and latitudinal distance of intelligent driving. Our results describe that the intelligent vehicle does have the basic ability of cognizance, and it can make correct reactions corresponding to the driving environment, but its intelligence still needs improving.

Keywords: Driving behavior, cognitive science, pattern recognition, intelligent vehicle.

1 Introduction

Studying on intelligent vehicles with the abilities of cognizing driving environment and making decisions autonomously, based on theories about audio-visual information processing and key technologies on audio-visual collaborative algorithms, can greatly enhance our research strength in this field. Road traffic safety is restricted by man-vehicle-road (environment) [1], and the latest news shows that human factors give rise to 92% of all road accidents, up to 88.29% of which are caused by drivers, so we can see driver plays an important role in the road traffic control system. There are some advantages and disadvantages in this system, the former is the driver can rely on his own audio-visual features unconditionally, and cognize the surrounding traffic environment timely and effectively to make decisions; the latter is some invisible and uncontrollable factors from the driver, such as negligence, fatigue and so on, may lead to a large number of traffic accidents indeed. Therefore, to solve this huge security problem thoroughly, why not let the driver "walk" out of the road traffic control system, that is, the intelligent drive closed-loop system "vehicle - road (environment)" is created so as to improve security

and efficiency. The core of this new traffic control system is to get artificial intelligence theories and technologies involved, and make unmanned vehicles drive smoothly and intelligently in the future. Unmanned intelligent vehicles are able to learn the correct behaviors of human drivers, and discard the bad behaviors, furthermore there must be some day unmanned intelligent vehicles can drive safely and effectively as well as the human drivers, perhaps even better, so it is of great help to improve road safety significantly.

This paper focuses on the information processing of intelligent driving. First of all, on basis of cognitive science, we analyze the process of driving behaviors: beginning with the perception of driving environment, then matching the existed pattern in Memory, then making decisions to take corresponding actions; second, in accordance with construction perception and pattern recognition theory, further analysis is carried on; finally, using test data, we attempt to recover the exact intelligent behavior of the test vehicle by the characteristics of the latitudinal distance and driving velocity, and we can see the test vehicle has its own sensing and cognizing skills, which sets up a good foundation of our research on intelligent vehicles in the future.

2 Overview of Cognitive Psychology

Information processing is the core of cognitive psychology. Compared with computer, human brain is considered as the information processing system according to information processing theory. The major areas of cognitive psychology include perception, attention, representation, learning and memory, thinking and speech, and other psychological processes or cognitive processes, also pattern recognition and knowledge of the organization. The core is to reveal the internal psychological mechanisms of cognitive processing, that is, how to obtain, storage, process and use information.

There are two opposite points of views in perception theory: the first one is (i.e. construction perception) that human's perception depends on knowledge and experience, people base on the existing knowledge of things in the natural environment, and are guided by the relationship between the perception of activities and the natural knowledge; the second view denies the existing knowledge and experience, but holds that the perceptive stimulation is complete, also can provide rich information, which people can make use of to generate corresponding perceptual experience, instead of relying on the past experience. However, generally speaking, cognitive psychologists have enough reasons to believe perception depends on the existing knowledge and experience, and perceptive information is the result of the interaction between real stimulation information and memory information [2]-[4].

3 Research on Intelligent Driving

Intelligent driving, relying on in-vehicle intelligent devices with computer systems[5], develops driver's visual and auditory performances, real-timely interacts with the driving environment and perceives the environment around the vehicle using some sensors, such as laser radar, ultrasonic sensors, microwave radar, Global Position System, odometer and magnetic compass etc., then according to the road, vehicles

and obstacles autonomously controls steering and speed, intelligently plans their own behaviors, so that the intelligent vehicles can safely and reliably run on the road [6].

3.1 Information Process for Intelligent Driving Behavior

As shown in Fig. 1, human receive information by visual, auditory, tactile, and other sensors, then the information from these sensors will be transmitted via the bus and protocol into Electronic Control Unit (ECU). Inside ECU, the input information will be identified to make appropriate decisions, and Memory is needed in all these procedures. Finally, the information processing system outputs information to take appropriate actions issued by the Controller.

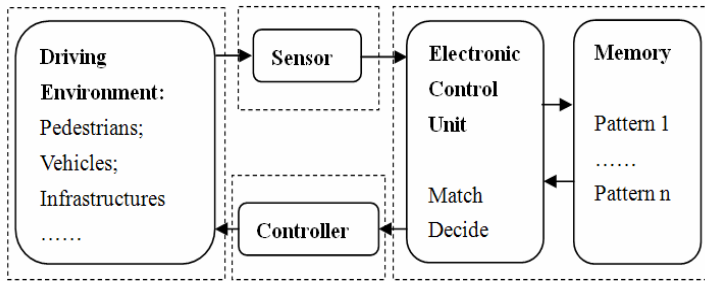


Fig. 1. Information Processing of Intelligent Driving Behavior

The process of intelligent driving behavior includes cognitive phase, decision phase and operation phase, so a variety of vehicle behaviors will be acquired through these three phases, using all kinds of equipments, computers to replace human brain, completing cognizing and decision-making phases.

3.2 Model of Intelligent Driving Behavior Based on Construction Perception

Cognitive psychologists generally confirm that perception depends on the past experiences and knowledge, and perception is the result of the interactions between stimulating information and memory, which is the main idea of construction perception (i.e. hypothesis test theory).

Construction perception holds the idea that the knowledge and experiences in the past play a significant part in assumptions, expectations in perception. People base on the past experiences to form the assumption of the current stimulus. Bruner (1957) and Gregory (1970) considered that perception was a construct procedure including hypothesis and test [7]. First people receive information, make assumption and test assumption, second, accept or search for information again, then test this assumption, until validate this assumption, thus they can make the correct explanation of this stimulus[8].

Based on construction perception, comparing and analyzing the detailed recognition process of sensing stimulation, receiving information and perception provides a sufficient basis to the following research.

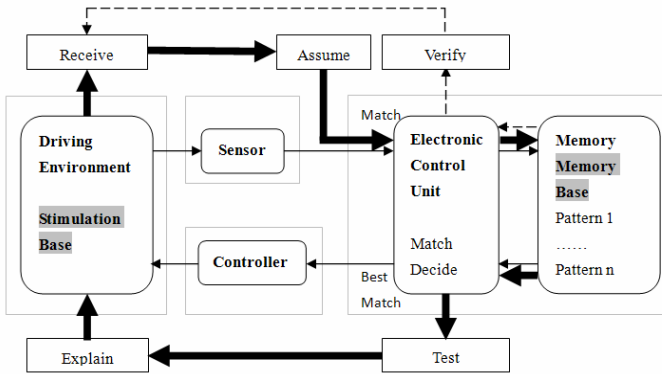


Fig. 2. Information Processing Model of Intelligent Driving Behavior Based on Construction Perception

Fig. 2 tells the information processing model of intelligent driving behavior based on construction perception and the bold arrows show the core of construction perception. Receiving information, making and testing assumptions again and again, help to validate an assumption, and give correct explanation of sensory stimulus.

3.3 Model of Intelligent Driving Behavior Based on Pattern Recognition

Pattern refers to a certain number of elements or components to form a certain stimulus obeying some rules. Pattern recognition is to detect, identify and confirm patterns, that is, through the human brain to match the stimulus previously acquired and stored in the long-term memory with the current information, to convert and analyze stimulus information for identification of external things, in order to include or expand into the cognitive structure of the human brain.

Prototype pattern recognition regards some abstract model existed in human's long-term memory as a prototype, which contains many similar shapes of the same pattern, obtaining identification by matching. In prototype pattern recognition, the external stimulus only compares with the prototype, because the prototype is a general characterization, this comparison only needs approximate matching.

In accordance with the views of the prototype matching model, the external stimulus has to compare with the prototype, not the accurate match, but the approximate match. This means as long as the existence of the corresponding prototype, the new, unfamiliar patterns should be identified to make the pattern recognition more flexible and adapt to environment [9]. Based on prototype pattern recognition of cognitive psychology theory, all elements in driving environment, such as weather, pedestrians, and road alignment etc., can stimulate and affect driving behaviors. We name these different elements a stimulation base, and then sum up different driving behaviors in different environments, such as obstacle avoidance behavior, turning behavior etc..

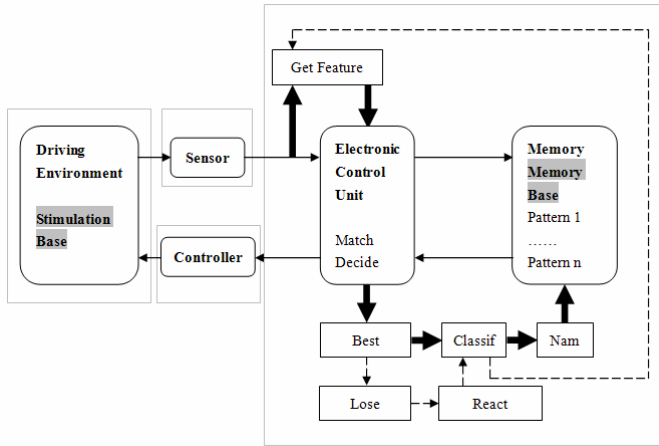


Fig. 3. Pattern Recognition Model of Intelligent Driving Behavior Based on Prototype Matching

Fig. 3 is a prototype pattern recognition model of intelligent driving behavior. Extracting features of the received information, ECU commands the knowledge base to match decisions, and then an exact match is created. After classifying and naming, the results enter the knowledge base, if not succeed, we need to re-feature extraction, and then match; if the accurate matching result loses characteristics, we need to tend to react, and then classified, and eventually into the knowledge base, and output an implementation after a successful match.

4 Analysis of Intelligent Driving Behavior

On basis of "Future Challenge" contest, we chose Vehicle A to make analysis of intelligent driving behaviors. Vehicle A sensed the leading car speed decreasing, and also the forward distance, in view of the construction perception model, first of all this stimulus tended to form Assumption 1, then passed into ECU, which controlled the Assumption 1 to match the patterns in Memory one by one, if not succeed then return to ECU, and tested Assumption 1 again, got stimulating information for the second time, formed Assumption 2, through ECU matched in Memory, this time if succeed, Assumption 2 was verified, the correct implementation would be taken by Controller, that was to change lane, overtake and return to the original lane. Moreover, according to the pattern recognition model based on the prototype, the message features that the leading car speed reducing and the forward distance lessening was perceived and obtained by intelligent Vehicle A, following the best match, classified, named and stored in Memory, a new pattern was created. If the feature was missing after the match, reaction tendency would be classified and returned to regain features.

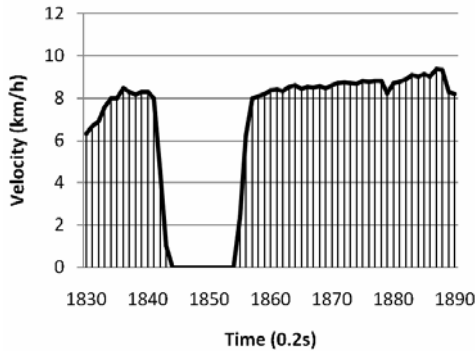


Fig. 4. Velocity of Intelligent Drive in “Future Challenge”

In Fig. 4, the result suggested the velocity changes of Vehicle A during the process of overtaking, and we could find in 1840th to 1855th (the area between two arrows), there was a 2-second period of 0 speed which tells Vehicle A did perceive and recognize the forward moving obstacle getting bigger and closer while running, so it stopped to take some time to make decisions. It was called reacting time, about 2 seconds, obviously it did need a few seconds that were much more than human beings’, but it implied the great ability of perceiving and decision-making, we believe that it must be improving in the future, which lays a solid foundation for the future research. Since then, Vehicle A started to overtake, shown in 1855th to 1885th (the area between the two dots), first speeded up to a stable value, about 9 km/h, and then fell to the original speed of about 8 km/h after passing.

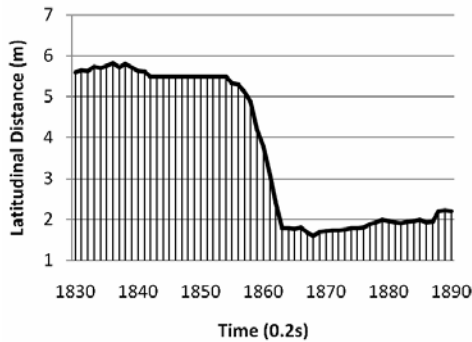


Fig. 5. Latitudinal Distance of Intelligent Drive in “Future Challenge”

Latitudinal distance changing could be found in Fig. 5, corresponding to its reacting time for perceiving and decision-making when stopped, a constant latitudinal distance (the area between two arrows) was seen during this period, so absolutely the area following this one was the overtaking process (the area between two dots). The curve trend showed, the latitudinal distance in the process of passing the leading car decreased a bit, nearly from 5.5 meters down to about 2 meters, and maintained

around 2 meters, so we could infer Vehicle A failed to return to the original lane after overtaking, it was probably because after overtaking there were no signs or identifiable objects in front of it.

We know that intelligent vehicles have autonomous driving capacities, such as keeping running in the same lane, perceiving forward objects, correctly identifying driving environment, safely avoiding obstacles and so on. In this overtaking behavior tests, however, intelligent driving abilities still need improving, for example, shirking the reacting time, increasing running speed much safely, and going back to the lane after overtaking and so on.

5 Results and Discussions

This article gives a new way of perception based on information processing for control system in intelligent vehicles, on the basis of construction perception and prototype pattern recognition theory. The results of overtaking behavior indicate the current intelligent driving ability in China, and the main conclusions are as follows:

(1) On theoretical basis of cognitive psychology, using construction perception (hypothesis test theory) we have established intelligent driving behavior model, clarified the short decision-making process, reflecting the assuming and testing again and again during decision-making, to get the best match;

(2) Intelligent driving behavior model based on prototype pattern recognition theory has been built in order to well match the stimulus and decision pattern;

(3) For the overtaking process, respective velocity and latitudinal distance graphs have been analyzed. During overtaking, intelligent Vehicle A slowed down first and stopped to perceive and make decisions, and then accelerated to overtake, then decreased to normal speed; the latitudinal distance was getting shorter and was kept in 2 meters with the reason that it failed to return the original lane;

(4) As a whole, intelligent Vehicle A has the basic cognitive ability, can not only correctly identify the driving environment, but also perceive the movement of the forward vehicle, and change lanes autonomously, avoid obstacles safely, keep running in the same lane and other intelligent driving behaviors, but efficiency and autonomy in intelligent driving will be better improved in the future

However our results have limitations, because all the tests were taken in ideal situations, without influences from other complicated environmental elements. Meanwhile, it should be noted that this study has examined only the overtaking behavior and further studies on other intelligent driving behavior will be summarized in our next paper.

Acknowledgments. The financial support from the National Natural Science Foundation of China, major project “Research on Comprehensive Test Environment Design and Evaluation Architecture of Intelligent Behavior for Unmanned Ground Vehicles” under Contract No. 90920304 was gratefully acknowledged and sincere gratitude was also extended to all participants for their valuable efforts in completing this study.

References

1. Parent, M., Yang, M.: Road Map towards Full Driving Automation. In: Proceedings of 8th International Conference on Application of Advanced Technologies in Transportation Engineering, Reston, pp. 663–668 (2004)
2. Zhang, Y.: Human-computer interface design based on knowledge of cognitive psychology. *Computer Engineering and Application* 30, 105–139 (2005)
3. Xiong, Z.: Systematic description for classical style of cognitive psychology. *Psychology Science* 6, 1510–1511 (2005)
4. Solso Robert, L., MacLin Kimberly, M., MacLin Otto, H.: *Cognitive Psychology*, Beijing (2005)
5. Yang, M.: Research on Unmanned Autonomous Driving Vehicles. *Journal of Harbin Institute of Technology*, 1259–1262 (2006)
6. Qin, S., Liu, J., Wen, W.: Research on Autonomous Intelligent Driving Behavior Model. *Journal of Beijing University of Aeronautics and Astronautics*, 793–796 (2003)
7. Zhou, H., Fu, X.: Cognitive science: the scientific frontier of new millennium. *Journal of Psychology Science Prospect*, 388–397 (2005)
8. Wang, G., Wang, A.: *Cognitive Psychology*, Beijing (2001)
9. Wilson, R.A., Keil, F.C.: *The MIT Encyclopedia of the Cognitive Sciences* (2000)

Inhibition of Return and Stroop Effect in Spatial Attention

Jia Fu

School of Business, Sun Yat-sen University,
510275 Guang Zhou, China
psy.jiajia@yahoo.com.cn

Abstract. In this study we examine the level at which inhibition of return (IOR) affects the processing of visual spatial stimuli. Experiment examined the effect of IOR on spatial Stroop interference. This results shows IOR could reverse the Stroop effects. We suggest that when attention is drawn away from a location, there is temporary inhibitory tagging of stimuli that are presented there. This tagging extends to the semantic and response-relevant properties of stimuli, helping to bias attention away from old and towards new events.

Keywords: Inhibition of return, inhibitory tagging, spatial Stroop effect.

1 Introduction

Human visual attention allows people to input from a large selection of visual information relevant to the current behavior, and inhibition of irrelevant stimuli, thus inhibiting is an important. Inhibition of return (IOR) refers to the observation that the speed and accuracy with which an object is detected are first briefly enhanced (for perhaps 100-300 milliseconds) after the object is attended, and then detection speed and accuracy are impaired (for perhaps 500-3000 milliseconds). It has been suggested that IOR promotes exploration of new, previously unattended objects in the scene during visual search or foraging by preventing attention from returning to already-attended objects. IOR is usually measured with a cue-response paradigm, in which a person presses a button when s/he detects a target stimulus following the presentation of a cue that indicates the location in which the target will appear. IOR was first described by Michael Posner and Yoav Cohen, who discovered that, contrary to their expectations, reaction time (RT) was longer to detect objects appearing in previously cued locations. It was subsequently shown that IOR can also be associated with a previously attended object, and under appropriate conditions these two inhibitory effects appear to be additive.

The mechanism on the IOR has long been in dispute. Fuentes et.al (1999) showed that stimulus processing is affected when stimuli are presented to locations subject to inhibition of return. They argued that activated representations of stimuli presented at inhibited locations are disconnected from their associated responses through an "inhibitory tagging" mechanism occurring in inhibition of return. Vivas et al (2001,2003,2006) presented the Stroop stimuli in locations subject to IOR (cued

locations) or in noninhibited (uncued) locations. The Stroop interference was reduced when stimuli fell at cued locations. This reduction was mainly due to the fact that congruent and neutral targets showed an increase in RTs when presented at cued compared with uncued locations (IOR effects). However, incongruent targets did not show such an increase; that is, IOR was not observed in this condition. This pattern of results is compatible with the idea of an inhibitory tagging mechanism that inhibits the prepotent tendency to read the word in the Stroop task when stimuli fall at cued locations. The fMRI study showed that in long SOA conditions (generated IOR condition) compared the short SOA condition (facilitation condition), spatial cues activate the bilateral anterior parietal cortex, but in color was not found in the IOR. And only found in the color IOR activation of the prefrontal cortex, both of which can explain non-spatial IOR and spatial IOR have different processing mechanisms (Chen, 2006). Then, in the cue-target paradigm, attention to spatial information processing and non-spatial information processing may also be different. However, Funes and Lupiáñez (2005, 2008) showed that spatial Stroop effect reduce in cued location, but no found IOR. Why two findings are inconsistent? The present experiment was designed mainly to investigate whether and how IOR will affect spatial Stroop effects. To achieve this goal: stimuli were presented either in up or below of the visual field, this avoid Simon effect affect the measurement of Stoop effect.

2 Method

2.1 Participants

Fourteen naïve, paid volunteers and the first author participated in Experiment . All participants (6 female, 8male, 21.8 ± 3.6 years) had normal or corrected-to-normal vision and gave written informed consent. The local ethics committee approved this and all the following experiments.

2.2 Apparatus and Stimuli

Participants were seated in a dimly lit, sound attenuated room and sat about 57 cm from an 85-Hz CRT monitor (iiyama MA203DT, resolution 1024×768) using a chin rest. All stimuli were presented on a black background.. An IBM compatible Pentium-equipped PC running Experiment-Builder 1.2 controlled stimulus presentation. A standard nine-point grid calibration was performed at the beginning of each block and the calibration was checked at the beginning of each trial. Participants self initiated each trial by pressing a defined button on a gamepad. Each square box was 3.3 degrees on a side and separated from its near neighbour by 3.3 degrees (Taylor, 2005; Funes, 2007).

2.3 Procedure

Fig. 1 illustrates the procedure. At the beginning of each trial, the fixation point and the appeared for 750ms.and then the cue appeared for 100ms.Following the cue, the fixation point and peripheral boxes remained on the screen for 200ms.The target was

then presented inside one of the boxes for 75ms. The fixation point and peripheral boxes remained alone on the screen until the participant's response, or for 1500ms if no response was made. Then, the next trial began. Auditory feedback was provided when an incorrect response was made.

Participants were instructed to press the X key on the keyboard with their left index finger when the arrow pointed to the top and to press the M key with their right index finger when the arrow pointed to the bottom. Peripheral cues appeared equally often to the up and bottom of fixation and were not informative about target location.

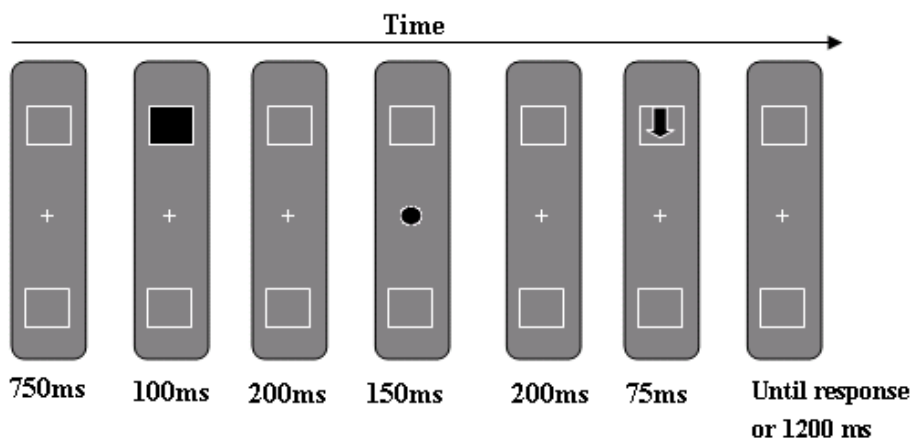


Fig. 1. Schematic view of a trial sequence

2.4 Design

The experiment was a two-factor (cueing \times spatial congruency) within-participant design. The factor cueing had two levels: valid trials in which the target and cue appeared at the same location, and invalid trials in which they appeared at the opposite locations. The spatial congruency variable also had two levels: On congruent trials, the direction in which the arrow pointed was consistent with its actual location. On incongruent trials, the arrow's direction was opposite to its actual location. The experiment consisted of five blocks of 240 trials. Each block comprised of six repetitions of factorial combination of two possible cue locations, two possible target locations, and two types of. All trials within each block were presented randomly. Prior to the experimental trials, a block of 40 trials was presented for practice. The total duration of the experiment was approximately 40 min with time for breaks between the blocks.

2.5 Result

The RTs above 1,200 ms or below 100 ms were discarded from the data analyses. Less than 6% of the trials were discarded following that criterion. Fig 2 shows the mean RTs. Within-participant design analysis of variance (ANOVA) was conducted to analyze both mean reaction times (RTs) and error percentages. The ANOVA

included cuing (valid vs invalid) and spatial congruency (congruent vs incongruent). The main effect of Cue Validity significant, $F(1, 12) = 14.34$, $p < 0.05$; significant main effect of spatial congruency, $F(1, 12) = 43.76$, $p < 0.001$, indicating a consistent response under conditions to be faster than inconsistent responses. Importantly, the effectiveness and spatial congruency leads significant interaction, $F(1, 12) = 22.56$, $p < 0.001$.

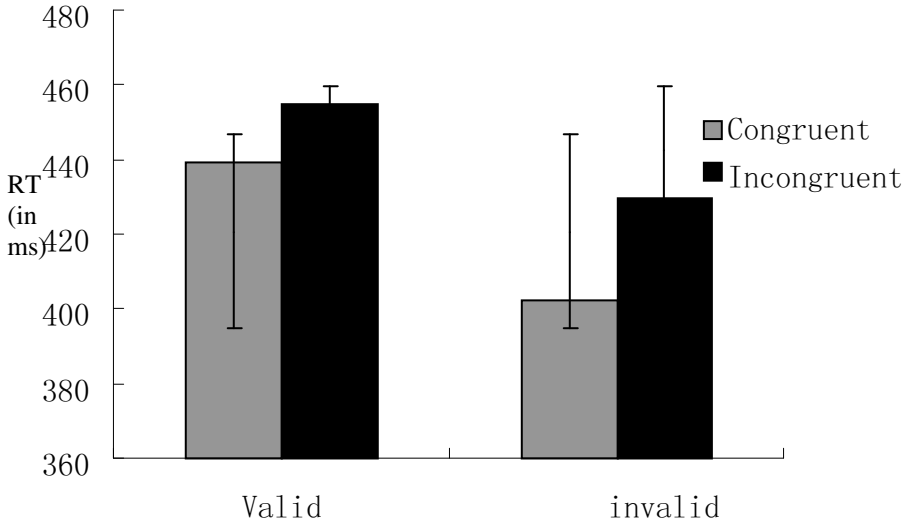


Fig. 2. Mean reaction times (RTs) for valid and invalid conditions

There was significant interaction between cuing effect and spatial congruency and further analysis of simple effects: Consistent and inconsistent conditions in under the simple effect of cue validity significantly, $F(1, 12) = 46.23$, $p < 0.01$; $F(1, 12) = 18.57$, $p < 0.05$, shows that subjects appear in the trail of the position of the stimuli appears to be slower than non-cued location in the stimulus, resulting in IOR. Cues to effective conditions, the spatial congruency of the simple effect was not significant $F < 1$, and invalid cue conditions, the spatial congruency of a significant simple effect $F(1, 12) = 40.99$, $p < 0.001$.

3 General Discussion

In current Experiment, we have found when central cue was showed, IOR regulate spatial Stroop, and in the cued location the amount of spatial Stroop is less than that in non-cued location the amount of spatial Stroop, that is, inhibition of return could play a role in conflict resolution. Funes (2007) no found inhibitory effect on the spatial Stroop impact. In their experiments, although the IOR does not work, but still found Stroop in the cued location to the decrease, due to the spatial congruency between stimulus and stimulus may require additional perception efforts to resolve the conflict.

This perception may be delayed to prompt attention from the IOR location and subsequent re-orientation. According to our experiment, in Funes's experiment, is not purely spatial Stroop, but joint Simon effect and spatial Stroop task. Although Fuens (2005), they have tried to separate spatial Stroop and Simon effect, their experiments using up, bottom, left and right arrows show the four directions to examine the impact of the conflict, but this setting and is not successful, because even in the spatial orientation of the up and down arrows appear to point around the target, there will Simon effects still emerged, and when in horizontal direction, there were the spatial Stroop effect and Simon effect.

Lupianez (2005, 2008) research suggests that, when the spatial Stroop conflict or the perception of stimuli in the early identification of processing stages, spatial cues can reduce the interference conflict. Spatial cues only affect spatial Stroop, without affecting the Simon effect. And spatial Stroop reduced in cued location, which happened in perceptual processing stages. In fact Stroop effect can occur at multiple stages of processing—at the perceptual stage or at the semantics processing stage and the response stage. In addition, although both the Simon and spatial Stroop tasks share many similar properties (e.g., the location of the stimulus being the task irrelevant dimension), they have been studied in distinct manners to examine different theoretical issues. These two tasks are caused by different types of conflict (stimulus–response conflict for the Simon effect and stimulus–stimulus conflict for the Stroop effect) (Xu, 2004).

In conclusion, when the central cue present, spatial Stroop reduced in cued location was due to the role of inhibitory tagging, and when there is no central cue, spatial Stroop was decreased for event cue–target integration processes.

Acknowledgments. We thank the anonymous reviewers for their helpful comments and suggestions on an early version of this paper.

References

1. Abrams, R.A., Dobkin, R.S.: Inhibition of return: effects of attentional cuing on eye movement latencies. *Journal of Experimental Psychology: Human Perception and Performance* 20, 467–477 (1994)
2. Callejas, A., Lupiáñez, J., Funes, M.J., et al.: Modulations among the alerting, orienting and executive control networks. *Experimental Brain Research* 167(1), 27–37 (2005)
3. Chen, Q., Wei, P., Zhou, X.: Distinct neural correlates for resolving Stroop conflict at inhibited and noninhibited locations in inhibition of return. *Journal of Cognitive Neuroscience* 18(11), 1937–1946 (2006)
4. Fan, J., Gu, X., Guise, K.G., et al.: Testing the behavioral interaction and integration of attentional networks. *Brain and Cognition* 70(2), 209–220 (2009)
5. Fan, J., Posner, M.: Human attentional networks. *Psychiatrische Praxis, Supplement* 31(2), S10–S14 (2004)
6. Fuentes, L.J., Boucart, M., Alvarez, R., et al.: Inhibitory processing in visuospatial attention in healthy adults and schizophrenic patients. *Schizophrenia Research* 40(1), 75–80 (1999)
7. Fuentes, L.J., Campoy, G.: The time course of alerting effect over orienting in the attention network test. *Experimental Brain Research* 185(4), 667–672 (2008)

8. Funes, M.J., Lupiáñez, J., Milliken, B.: Separate Mechanisms Recruited by Exogenous and Endogenous Spatial Cues: Evidence From a Spatial Stroop Paradigm. *Journal of Experimental Psychology: Human Perception and Performance* 33(2), 348–362 (2007)
9. Funes, M.J., Lupiáñez, J., Milliken, B.: The modulation of exogenous Spatial Cueing on Spatial Stroop interference: Evidence of a set for cue-target event segregation. *Psicologica* 29(1), 65–95 (2008)
10. Fuentes, L.J., Santiago, E.: Spatial and semantic inhibitory processing in schizophrenia. *Neuropsychology* 13(2), 259–270 (1999)
11. Fuentes, L.J., Vivas, A.B., Alvarez, R., et al.: Inhibitory tagging in inhibition of return is affected in schizophrenia: Evidence from the stroop task. *Neuropsychology* 14(1), 134–140 (2000)
12. Fuentes, L.J., Vivas, A.B., Humphreys, G.W.: Inhibitory tagging of stimulus properties in inhibition of return: Effects on semantic priming and flanker interference. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology* 52(1), 149–164 (1999)
13. Fuentes, L.J., Vivas, A.B., Humphreys, G.W.: Inhibitory mechanisms of attentional networks: Spatial and semantic inhibitory processing. *Journal of Experimental Psychology: Human Perception and Performance* 25(4), 1114–1126 (1999)
14. Francesca, F., Massimo, S., Sandro, R., et al.: Determining priority between attentional and referential-coding sources of the Simon effect through optokinetic stimulation. *Neuropsychologia* 48(4), 1011–1015 (2010)
15. Lupiáñez, J., Klein, R.M., Paolo, B.: Inhibition of return: Twenty years after. *Cognitive Neuropsychology* 23(7), 1003–1014 (2006)
16. Lupiáñez, J., Funes, M.J.: Peripheral spatial cues modulate spatial congruency effects: Analysing the locus of the cueing modulation. *European Journal of Cognitive Psychology* 17(5), 727–752 (2005)
17. Lupiáñez, J., Ruz, M., et al.: The manifestation of attentional capture: facilitation or IOR depending on task demands. *Psychol. Res.* 71(1), 77–91 (2007)
18. Prime, D.J., Ward, L.M.: Inhibition of Return From Stimulus to Response. *Psychological Science* 15(4), 272–276 (2004)
19. Taylor, T.L., Ivanoff, J.: Inhibition of return and repetition priming effects in localization and discrimination tasks. *Canadian Journal of Experimental Psychology* 59(2), 75–89 (2005)
20. Vivas, A.B., Fuentes, L.J., Estevez, A.F., et al.: Inhibitory tagging in inhibition of return: Evidence from flanker interference with multiple distractor features. *Psychonomic Bulletin and Review* 14(2), 320–326 (2007)
21. Vivas, A.B., Fuentes, L.J.: Stroop interference is affected in inhibition of return. *Psychonomic Bulletin and Review* 8(2), 315–323 (2001)
22. Vivas, A.B., Humphreys, G.W., Fuentes, L.J.: Object-Based Inhibition of Return in Patients With Posterior Parietal Damage. *Neuropsychology* 22(2), 169–176 (2008)
23. Vivas, A.B., Humphreys, G.W., Fuentes, L.J.: Inhibitory processing following damage to the parietal lobe. *Neuropsychologia* 41(11), 1531–1540 (2003)

Two Dimensions Data Slicing Algorithm, a New Approach in Mining Rules of Literature in Traditional Chinese Medicine

Guang Zheng^{1,2}, Hongtao Guo^{2,3}, Yuming Guo^{2,3}, Xiaojuan He²,
Zhongxian Li¹, and Aiping Lu^{2,*}

¹ School of Information Science-Engineering, Lanzhou University,
Lanzhou, 730000, China

² Institute of Basic Research in Clinical Medicine, China Academy of Chinese
Medical Sciences, Dongzhimen, Beijing, 100700, China

³ Shanghai University of Traditional Chinese Medicine, Shanghai, 201203, China
lap64067611@126.com

Abstract. Chinese medicinal herbs, acupuncture, and patterns are important components in traditional Chinese medicine. In this paper, focused on facial nerve paralysis, we proposed an algorithm named two dimensions data slicing to mine rules of Chinese medicinal herbs, acupuncture, and patterns. The process of mining was done in two dimensions. The one-dimension analyzes the frequencies of Chinese medicinal herbs, acupoints, and patterns. The two-dimension analyzes the frequencies of co-existed keyword pairs. By examining the results of these two dimensions, although some noises existed, most regular knowledge of treating this disease is mined out. This algorithm might be useful in mining rules in the literature of traditional Chinese medicine.

Keywords: Data mining, facial nerve paralysis, two dimension data slicing algorithm, traditional Chinese medicine

1 Introduction

Acupuncture [22, 25, 26], Chinese medicinal herbs [21], and diagnoses [23] are important component procedures in traditional Chinese medicine which prevent and treat diseases. Since they are effective, simple in clinical practice, and low in cost, they have been widely used in China and countries around China for thousands of years.

The application of acupuncture and moxibustion through stimulating certain acupoints on the human body is to activate the meridians and collaterals so as to regulate the functions of the internal organs, qi, and blood so as to prevent and treat diseases [22].

Nowadays, there are many debates in the field of acupuncture [15, 16, 27]. However, in the clinical practice of traditional Chinese medicine, acupuncture is widely used in the treatment of facial nerve paralysis [24, 5, 11, 8, 9, 13, 14].

* Corresponding author.

Chinese medicinal herbs is one of the most important modalities utilized in traditional Chinese medicine (TCM). Each herbal medicine prescription is a cocktail of many herbs tailored to the individual patient. One batch of herbs is typically decocted twice over the course of one hour. The practitioner usually designs a remedy using one or two main ingredients that target the illness. Then the practitioner adds many other ingredients to adjust the formula to the patient's patterns. Sometimes, ingredients are needed as it is believed that it will cancel out toxicity or side-effects of the main ingredients. Some herbs require the use of other ingredients as catalyst or else the brew is considered to be ineffective. The latter steps require great experience and knowledge, and make the difference between what is accepted as a good Chinese herbal doctor and an amateur. Unlike western medications, the balance and interaction of all the ingredients are considered more important than the effect of individual ingredients. A key to success in TCM is the treatment of each patient as an individual [18].

For diagnoses in traditional Chinese medicine, there are four diagnostic methods to observe and diagnose disease. These methods are inspection, auscultation, inquiry, and pulse-taking and palpation, refer to the four basic procedures used in diagnosing a disease. They are the presuppositions of correct differentiation and effective treatment in traditional Chinese medicine [23].

In traditional Chinese medicine, acupuncturists believe that hand and foot yang meridians are all pass on to the top of face. If face meridians are blocked by external pathogen, especially meridian sinew of Small Intestine and Stomach Meridian dysfunction can lead to the occurrence of facial nerve paralysis [26,22].

In this paper, by examining data/text mining algorithms existed nowadays [7, 11, 2, 3, 6, 10, 12, 20], we created a two dimension data slicing algorithm. By executing it, we calculated the acupoint sets in the treatment of facial nerve paralysis in two dimensions. One-dimension is to calculate the frequency of single acupoints, Chinese medicinal herbs, and patterns existed in the literature. Two-dimension is to calculate the frequencies of co-existed pairs of them. Then, slice them according to their frequencies and then, compare them with the textbook of acupuncturology for further analysis.

As there maybe confusion in the translation of acupoints and other terminologies in traditional Chinese medicine, we refer to WHO standards [17,19].

2 Material and Methods

2.1 Material: Data Preparation

In SinoMed (<http://sinomed.imicams.ac.cn/index.jsp>), on the default query strategy, we queried on the keyword of “facial nerve paralysis” in Chinese. As a result, we fetched a data set of 8,261 records on Dec. 29, 2010. These data set are focused on the treatments and reviews on facial nerve paralysis. Take acupoints for example, we want to get different levels of acupoints in the treatment of facial nerve paralysis by text mining. Chinese medicinal herbs and patterns can be treated in the same way.

When the data set is ready, we get it formatted from un-structured text file into structured database. By doing this, we get a record set of 459 records formatted from 8,261 papers. They are filtered out from keywords and abstracts by the lists of acupoints. They contain the paper ID and the associated acupoints, line by line. In order to simplify the description of algorithm, these data are assumed to be stored in “TABLE_I” with fields of “PMID (paper ID)” and “keyword (name of acupoint)”. In this paper, we focus on the distribution of acupoints in treating facial nerve paralysis.

When TABLE_I is constructed, we will analysis these acupoints in two different dimensions. In one-dimension, these acupoints will be calculated for their frequencies in literature. That is to say, the frequencies of the acupoints’ occurrence will be calculated and listed in the descendent order. In two-dimension, the binary relationships of these acupoints will be calculated. What’s more, the co-occurrent frequencies are also calculated and assigned to the binary relationships. That is to say, the frequencies of co-occurrent acupoint-pairs will be calculated and listed in descendent order.

2.2 Method: One-Dimension Data Slicing Algorithm

One-dimension analyzes the frequencies of the acupoints existed in literature focused on “facial nerve paralysis”. Based on the data in “TABLE_I”, “keyword” will be calculated for their frequencies. The “keyword”, together with its frequency, will be stored in “TABLE_II” with fields of “keyword” and “ps (for frequency)”.

To be more specifically, we list the algorithm of one dimension in Table I.

Thus, in “TABLE_II”, the frequencies of acupoints can be calculated in field “keyword” in the descendent order.

Table 1. Algorithm of One Dimension

```

USE TABLE_I
DO WHILE NOT EMPTY
GO TOP
  kw = keyword
  frqcy =frequency(kw) //calculate the frequency of kw in the whole list
  USE TABLE_II
    APPEND BLANK
    REPLACE keyword WITH kw, ps WITH frqcy
  USE TABLE_I
  DELETE ALL keyword = kw
ENDDO
USE TABLE_II
SORT ON ps DESCENDENT //keywords sorted on descending order of ps

```

2.3 Method: Two-Dimension Data Slicing Algorithm

Based on the principle of algorithm described in [7], Two-dimension analyzes the frequencies of co-occurrent acupoint-pairs existed in literature focused on “facial nerve paralysis”. Based on the data in “TABLE_I”, acupoints in field “keyword” will be calculated according to each “PMID”. That is, according to one “PMID”, every two different acupoints associated with it form a binary relationship (co-occurrent acupoint-pair). They are stored in “TABLE_III” with fields of “lh” (left hand of co-occurrent acupoint-pairs) and “rh” (right hand of co-occurrent acupoint-pairs). Then, the frequencies of the co-occurrence acupoint-pairs can be calculated and stored in “TABLE_IV” with fields of “lh”, “rh”, and “ps” (for the frequency of co-occurrent acupoint-pairs).

To be more specifically, we list the algorithm of two-dimension in Table 2.

Thus, in “TABLE_IV”, the frequencies of co-occurrent acupoint-pairs can be calculated in fields of “lh” and “rh” in the descendent order of “ps”.

Table 2. Algorithm of Two Dimension

```

USE TABLE_I
FOR each PMID
  k =Number_of_keyword(PMID)
  j = 1
  FOR keyword(i) (i =, 1, 2, ..., k)
    DO while j ≤ k
      keyword_Pair=keyword(i)+keyword(j)
      j = j + 1
      OUTPUT keyword_Pair INTO TABLE TABLE_III
    ENDDO
  j = 1
  ENDFOR
ENDFOR

USE TABLE_III
DO WHILE NOT EMPTY
GO TOP
  tmp_lh = lh
  tmp_rh = rh
  kw = lh + rh
  frqcy =frequency(kw) //calculate the frequency of lh + rh
  USE TABLE_IV
  APPEND BLANK
  REPLACE lh WITH tmp_lh, rh WITH tmp_rh, ps WITH frqcy
  USE TABLE_III
  DELETE ALL lh = tmp_lh AND rh = tmp_rh
ENDDO
USE TABLE_IV
SORT ON ps DESCENDENT //lh + rh sorted on descending order of ps

```

3 Results

3.1 One-Dimension

One-Dimension: Acupoints

Applying algorithm described in Table 1, we mined the data set of all the acupoints existed in literature. Then, we found the following list with threshold greater equal than 12. This list of acupoints are ordered by the frequency in the descendent way.

Table 3. Acupoints in the treatment of facial nerve paralysis calculated in one dimension with frequency ≥ 12

Acupoint	Chinese	ID	Frequency	Noises
Dicang	地仓	ST4	132	No
Zhongshu	中枢	GV7	129	Yes
Jiache	颊车	ST6	116	No
Yangbai	阳白	GB14	110	No
Hegu	合谷	LI4	91	No
Yifeng	翳风	TE17	87	No
Xiaguan	下关	ST7	81	No
Taiyang	太阳	-	72	No
Sibai	四白	ST2	47	No
Yingxiang	迎香	LI20	45	No
Cuanzhu	攒竹	BL2	44	No
Renzhong	人中	GV26	43	No
Zusanli	足三里	ST36	35	No
Fengchi	风池	GB20	34	No
Yuyao	鱼腰	-	33	No
Chengjiang	承浆	CV24	23	No
Sizhukong	丝竹空	TE23	18	No
Taichong	太冲	LR3	14	No
Quanliao	颧髎	SI18	12	No

Comparing this table with textbook [22], there are some differences between them.

1. All the main/kernal acupoints of treating facial nerve paralysis are included i.e., Dicang, Jiache, Yangbai, Hegu, Sibai, Cuanzhu, and Quanliao.
2. For subsidiary acupoints, most of them are included. For example, Yifeng, Fengchi, Zusanli, Shuigou, Yingxiang, Quchi, and Yuyao.
3. More interesting, there are also acupoints which are not included in textbook, yet commonly adapted in the clinical practise of acupuncture. They are Taiyang, Xiaguan, Sizhukong, Chengjiang, and Taichong. They might be a complementarity for the textbook, and we will discuss it in the later section.

One-Dimension: Chinese Medicinal Herbs

Applying algorithm described in Table 1, we mined the data set of all the Chinese medicinal herbs existed in literature, we found the following list with threshold greater equal than 10. This list of herbs is ordered by the frequency in the descendent way.

Table 4. Chinese medicinal herbs in the treatment of facial nerve paralysis calculated in one dimension with frequency ≥ 10

PinYin	Latin Name	Chinese	Frequency	Noise
Chuanxiong	Ligusticum chuanxiong Hort.	川芎	31	No
Fuzi	Radix Aconiti Lateralis Preparata	附子	30	Partially
Jiangcan	Bombyx Batryticatus	僵蚕	30	No
Quanxie	Scorpio	全蝎	28	No
Baifuzi	Rhizoma Typhonii	白附子	24	No
Fangfeng	Raidix Saposhnikoviae	防风	23	No
Danggui	Angelica sinensis	当归	20	No
Danshen	Radix Et Rhizoma Salviae Miltiorrhizae	丹参	20	No
Maqianzi	Semen Strychni	马钱子	17	No
Huangqi	Leguminosae	黄芪	15	No
Guizhi	Ramulus Cinnamomi	桂枝	13	No
Shengjiang	Rhizoma Zingiberis Rescens	生姜	12	No
Tianma	Rhizom Gastrodiae	天麻	11	No
Baizhi	Radix Angelicae Dahuricae	白芷	11	No
Chishao	Radix Paeoniae Rubra	赤芍	10	No

From Table 3, we found characteristics of Chinese medicinal herbs mined out in the treatment of facial paralysis. They compose the prescription of *Qianzheng San* (composed by Quanxie, Jiangchan, and Baifuzi) which is a famous formula in traditional Chinese medicine and widely used in clinical practice. As to other hers, i.e., in PinYin, “Baifuzi”, “Fuzi”, “Huangqi”, “Guizhi”, and “Shengjiang”, they are used for warming yang and dissipating cold; “Chuangxiong”, “Danggui”, “Danshen”, and “Chishao” are used for activating blood; “Jiangchan”, “Quanxie”, “Fangfeng”, “Tianma”, and “Baizhi” are used for dispelling wind to free the collateral vessels. As for toxic and acrid herbs, they are adopted for external use, such as “Maqianzi” and “Shengjiang”.

One Dimension: Patterns

Applying algorithm described in Table 1, we mined the data set of all the patterns existed in literature, we found the following list with threshold greater equal than 24. They are shown in Table 3.

In this table, it is clear that there are much variety in the column tagged “Noise”. By checking in the original data set of literature downloaded from SinoMed, we can see that “Tinnitus”, “Vomit”, “Dizziness” etc. that are tagged with “No” in “Noise” column. They are all the sensitive keywords of patterns yet not associated with facial nerve paralysis. As to “Leakage”, “Anthema”, and

Table 5. Patterns occurred in facial nerve paralysis calculated in one dimension with frequency ≥ 24

English Name	Chinese	Frequency	Noise
Herpes	疱疹	332	No
Paralysis	瘫痪	248	No
Leakage	漏	185	Partially
Facial distortion	口眼歪斜	171	No
Anthema	疹	171	Partially
Numbness	麻木	158	No
Headache	头痛	149	Partially
Tinnitus	耳鸣	111	Yes
Vomit	呕吐	75	Yes
Dizziness	头晕	60	Yes
Deaf	耳聋	60	Yes
Nausea	恶心	54	Yes
Otalgia	耳痛	51	Yes
Fever	发热	47	Yes
Twitch	抽搐	35	Yes
Sialorrhea	流涎	29	No
Ptosis	眼睑下垂	25	No
Hemiparalysis	半身不遂	24	Yes

“Headache” are tagged with “Partially”, by checking in the original data set, they are partially associated with facial nerve paralysis.

In this table, too many noises are shown. Take “Headache” for example, by checking it in the original literature, we found that it can be filtered out by associated with “spasm”, “non-headache”, “tumor”, “virus”, “cerebral apoplexy”, etc. However, by checking the keywords, these papers are tagged with facial nerve paralysis, that is, there is too much noise in the literature downloaded from SinoMed.

3.2 Two-Dimension

When the set of acupoints associated with its paper ID is ready, we applied the data slicing algorithm proposed in Table 2.

Two-Dimension: Acupoints

Applying algorithm described in Table 2, we mined the co-existed acupoint pairs, and the result is listed out by the order of frequency. When visualized, they are Fig. 1 and 2. In these figures, round nodes are connected with each other. Nodes are tagged with the name of acupoint. Numbers on the lines represents the frequencies of these two acupuncture point occurring together.

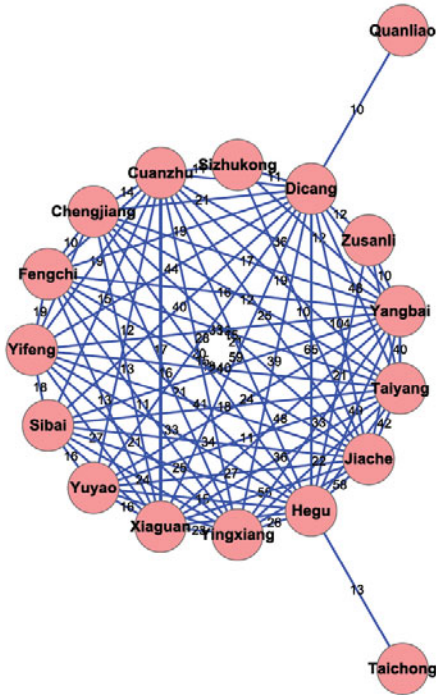


Fig. 1. Co-existed acupoint pairs with threshold ≥ 10

In Fig. 1, the threshold of frequency is 10 ($frequency \geq 10$). By this threshold, we get the acupoint set of “Yangbai”, “Sizhukong”, “Cuanzhu”, “Jiache”, “Hegu”, “Taiyang”, “Xiaguan”, “Yingxiang”, “Yuyao”, “Sibai”, “Yifeng”, “Fengchi”, and “Dichang”. This set indicates the points in the treatment of facial nerve paralysis. They are mainly on meridians of Large Intestine, Stomach, Small Intestine, and Bladder. That is, in the clinical practice of acupuncture, they covered the meridians of Small and Stomach as discussed before, but not limited within them.

In Fig. 2, the threshold is updated to 40. By this, we get the acupoint set of “Jiache”, “Dichang”, “Yangbai”, “Taiyang”, “Yifeng”, “Hegu”, “Xiaguan”, and “Sibai”. They are the most commonly used acupoints in the treatment of facial nerve paralysis. Meanwhile, on checking the distribution of these acupoints, we know that they are mainly located in the vicinity of disease sites. This shows acupuncture treatment of facial paralysis mainly depend on local and nearby therapeutic effect.

Two-dimension: Chinese Medicinal Herbs

Applying algorithm described in Table 2, we mined the co-existed Chinese medicinal herb pairs, and the result is listed out by the order of frequency. When

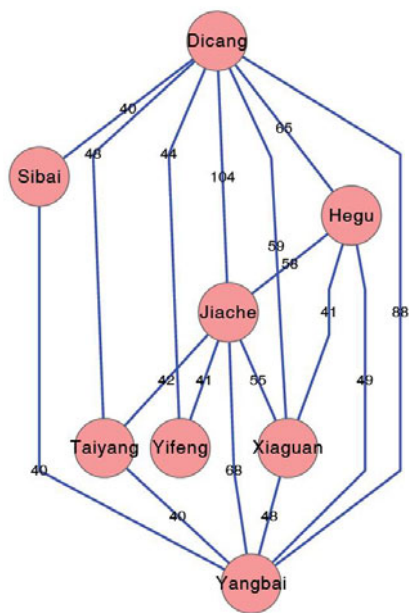


Fig. 2. Co-existed acupoint pairs with threshold ≥ 40

visualized, they are Fig. 3 and 4. In these figures, round nodes are connected with each other. Nodes are tagged with the name of herbs. Numbers on the lines represents the frequencies of these two herbs occurring together.

Fig. 3 Medication rule of facial nerve paralysis with threshold ≥ 15 . It shows that Bombyx Batryticatus (Jiangcan, 僵蚕), Rhizoma Typhonii (Baifuzi, 白附子), Rhizoma Chuanxiong (Chuanxiong, 川芎), Radix Saposhnikoviae (Fangfeng, 防风) and Scorpio (Quanxie, 全蝎) were commonly used in facial nerve paralysis. More detailed explanation can be found in the explanation of Table 1.

By this threshold, it shows that Stiff Bombyx Batryticatus, Rhizoma Typhonii, Rhizoma Chuanxiong, Radix Saposhnikoviae, Scorpio, and Radix Angelicae Sinensis (Danggui, 当归) are commonly used in the treatment of facial nerve paralysis.

The efficacy of Bombyx Batryticatus, Rhizoma Typhonii and Scorpio is to extinguish wind and resolve phlegm, free the collateral vessels and arrest convulsions. They are ingredients of fomular *Qianzheng San* which is widely used of facial nerve paralysis. Both of Radix Saposhnikoviae and Rhizoma chuanxiong as an important couplet medicinals play a role of dispelling wind and relieving pains. What's more, Rhizoma chuanxiong also has the efficacy of activating blood and moving qi. It corresponds to the pathogenesis of facial nerve paralysis in traditional Chinese medicine. On threshold 13, Radix Angelicae Sinensis can activate blood as well. But we can't see more herbs with the identical effect

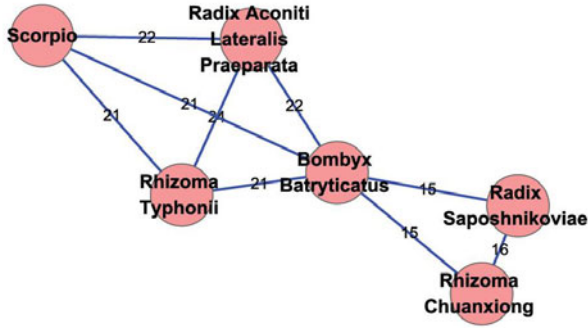


Fig. 3. Co-existed Chinese medicinal herb pairs with threshold ≥ 15

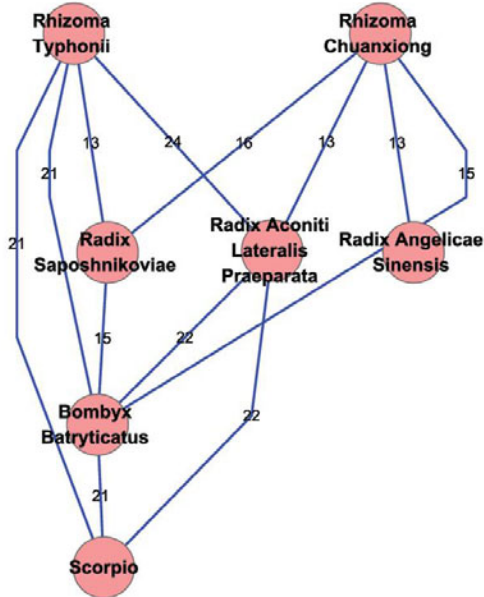


Fig. 4. Co-existed Chinese medicinal herb pairs with threshold ≥ 13

anymore in this threshold. In fact, more herbs with the function of activating blood and tonifying qi just like *Flos Carthami* and *Radix Astragali* are adopted in clinical practice. More than that, tonifying and harmonizing qi and blood conformed to the treatment principal of qi deficiency with blood stasis pattern in facial nerve paralysis. So we need to do further research on remitting the omission of general prescription and even on pattern differentiation of traditional Chinese medicine.

Two-Dimension: Patterns

Applying algorithm described in Table 2, we mined the co-existed pattern pairs, and the result is listed out by the order of frequency. When visualized, they are Fig. 5. In Fig. 5, round nodes are connected with each other. Nodes are tagged with the name of patterns. Numbers on the lines represents the frequencies of these two patterns occurring together.

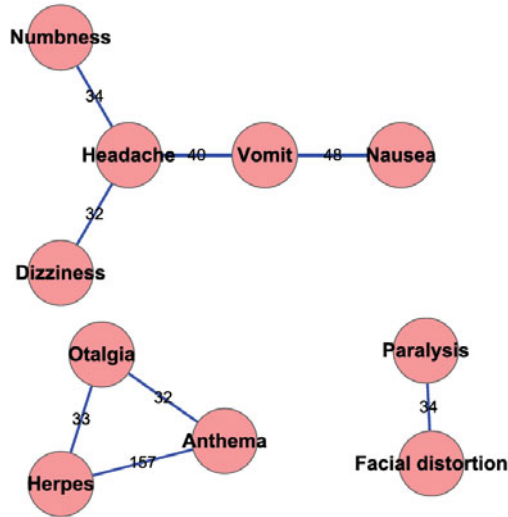


Fig. 5. Co-existed pattern pairs of facial nerve paralysis with frequency

In Fig. 5, patterns of “Headache-Vomit-Nausea” can be taken as noise according to the clinical manifestation of facial nerve paralysis. However, they are the clinical manifestation for disease of *Ramsay Hunt syndrome* [4]. This is caused by the over tagging of keywords in the literature of SinoMed. There are sub-networks are of the same kind, i.e., triangle sub-network of “Otagia-Herpes-Anthema” and sub-network of “Paralysis-Facial distortion”.

Although there are noise in the literature, we will develop other techniques for the further pre-treatment of the data set downloaded.

4 Conclusion and Discussion

Compared the data mining results of reviews and clinical observations [15, 16, 27, 24, 5, 11, 8, 9, 13, 14] with the textbooks [21, 23, 22], we found out that:

1. **Table 3 covers all the main points and most of the subsidiary acupoints.** This can be checked by comparing with this textbook of acupuncture [22], the high quality of mining result lies in the specifieness of the name of acupoints (at least most of them).
2. **Herbs mined out in Table 4 in the treatment of facial nerve paralysis are reasonable.** Compared with the theory of traditional Chinese medicines, most of the results are good. Of course, there is noise existed, i.e., “附子”(Radix Aconiti Lateralis Preparata) can be taken as the substring of “白附子”(Rhizoma Typhonii).
3. **Much items of noise existed in the mining result which can be seen in Table 5.** Compared with the patterns known for facial nerve paralysis, the result mined out in Table 3 is not good. By checking them in original data set, we found that the main reason is the low quality of tags associated with papers: too many tags with general meaning associated to literatures which is the main reason of noise existed in patterns. In order to tackle this problem, more rules based on professional knowledge should be constructed and applied in the future mining process.
4. **Based on the whole mining results, we can see that the antinoise is built on the specifieness of filters of sensitive keywords.** By Table 3, 4 and 5, we know that more specific the filters are, the more correct the mining results.
5. **The algorithm, named *two dimensions data slicing*, might be useful in mining the literature of traditional Chinese medicine.**

Acknowledgement. This work was partially supported by the National Eleventh Five Year Support Project of China (2006BAI04A10), the Innovative Methodology Project supported by MOST of China(2008IM020900). National Science Foundation of China (No. 30902003, 30973975, 90709007, and 81072982).

References

1. Razali, A.M., Ali, S.: Generating Treatment Plan in Medicine: A Data Mining Approach. *Shahriyah Ali Journal: American Journal of Applied Sciences* 6(2), 345–351 (2009)
2. Hotho, A., Nürnberger, A., Paaß, G.: A Brief Survey of Text Mining. *LDV Forum - GLDV Journal for Computational Linguistics and Language Technology* 20(1), 19–62 (2005)
3. Hand, D.J., Mannila, H., Smyth, P.: *Principles of Data Mining*. MIT Press, Cambridge (2001); ISBN 026208290X, 9780262082907

4. Lanska, D.J.: Ramsay Hunt syndrome. *J. Neurol Neurosurg Psychiatry* 71, 149–154 (2001); doi:10.1136/jnnp.71.2.149
5. Fu, X.: Observation on therapeutic effect of acupuncture on early peripheral facial paralysis. *Zhongguo Zhen Jiu* 27(7), 494–496 (2007)
6. Tzanis, G., Berberidis, C., Vlahavas, I.P.: Biological Data Mining. *Encyclopedia of Database Technologies and Applications*, 35–41 (2005)
7. Zheng, G., Jiang, M., Xu, Y., Chen, G., Lu, A.: Discrete Derivative Algorithm of Frequency Analysis in Data Mining for Commonly-existed Biological Networks. In: *CNMT 2010*, pp. 5–10 (2010)
8. He, X.W., Wang, B.Y., Huang, J.H., Zhou, S.L., Huang, X.B., Wu, J.: Clinical observation on moxibustion and acupuncture at Zusanli (ST 36) for treatment of refractory facial paralysis. *Zhongguo Zhen Jiu* 26(11), 775–777 (2006)
9. Lei, H., Wang, W., Huang, G.: Acupuncture benefits a pregnant patient who has Bell's palsy: a case study. *J. Altern. Complement. Med.* 16(9), 1011–1014 (2010)
10. Geng, L., Hamilton, H.J.: Choosing the Right Lens: Finding What is Interesting in Data Mining. *SCI*, vol. 43, pp. 3–24. Springer, Heidelberg (2007)
11. Rosted, P., Woolley, D.R.: Bell's Palsy following acupuncture treatment—a case report. *Acupunct. Med.* 25(1-2), 47–48 (2007)
12. Schmidt, S., Vuillermin, P., Jenner, B., Ren, Y., Li, G., Phoebe Chen, Y.-P.: Mining Medical Data: Bridging the Knowledge Divide. In: *Proceedings of eResearch Australasia* (2008)
13. Shen, T.L., Zhang, W., Li, Y.: Clinical observation on the principle of single usage of acupoints of Shaoyang meridian for treatment of facial paralysis in acute stage. *Zhongguo Zhen Jiu* 30(6), 461–464 (2010)
14. Song, X.J., Zhang, D.: Study on the manifestation of facial infrared thermography induced by acupuncture Guangming (GB 37) and Hegu (LI 4). *Zhongguo Zhen Jiu* 30(1), 51–54 (2010)
15. Suzuki, J., Kobayashi, T.: Internet information availability about facial nerve paralysis. *Nippon Jibiinkoka Gakkai Kaiho* 113(11), 844–850 (2010)
16. Tong, F.M., Chow, S.K., Chan, P.Y., Wong, A.K., Wan, S.S., Ng, R.K., Chan, G., Chan, W.S., Ng, A., Law, C.K.: A prospective randomised controlled study on efficacies of acupuncture and steroid in treatment of idiopathic peripheral facial paralysis. *Acupunct Med.* 27(4), 169–173 (2009)
17. WHO, WHO International Standard Terminologies on Traditional Medicine in the Western Pacific Region, WHO Library Cataloguing in Publication Data (March 2007); ISBN 978-92-9061-248-7
18. Wikipedia (2011), http://en.wikipedia.org/wiki/Chinese_herbology
19. World Health Organization Western Pacific Region, Acupuncture Point Location. In: *The Western Pacific Region*, People's Medical Publishing House (February 2010); ISBN 978-7-117-12332-7
20. Wu, X., Kumar, V., Ross Quinlan, J., et al.: Top 10 algorithms in data mining. *Knowl. Inf. Syst.* 14, 1–37 (2007)
21. Gao, X.: *Science of Chinese Pharmacology*. China Press of Traditional Chinese Medicine (February 2010); ISBN 978-7-80156-318-7
22. Shi, X.: *Acupuncturology*. China Press of Traditional Chinese Medicine (2008); ISBN: 978-7-80156-314-9
23. Zhu, W.: *Diagnostics of Traditional Chinese Medicine*. China Press of Traditional Chinese Medicine (February 2010); ISBN 978-7-80156-310-1

24. Zhang, D.: A method of selecting acupoints for acupuncture treatment of peripheral facial paralysis by thermography. *Am. J. Chin. Med.* 35(6), 967–975 (2007)
25. Li, Z., Zhang, Q.: *English Textbook for Traditional Chinese Medicine*. Shanghai Scientific and Technical Publishers (August 2009)
26. Zhou, Z., et al.: *Internal Medicine of Traditional Chinese Medicine*. China Press of Traditional Chinese Medicine (February 2007)
27. Zhou, M., He, L., Zhou, D., Wu, B., Li, N., Kong, S., Zhang, D., Li, Q., Yang, J., Zhang, X.: Acupuncture for Bell's palsy. *J. Altern. Complement. Med.* 15(7), 759–764 (2009)

Frequent Closed Pattern Mining Algorithm Based on COFI-Tree

Jihai Xiao¹, Xiaohong Cui¹, and Junjie Chen²

¹ College of Textile Engineering and Art, Taiyuan University of Technology,
030600 JinZhong, China

² College of Computer Science and Technology, Taiyuan University of Technology,
030024 TaiYuan, China

{xiaojihai,tycuixiaohong}@126.com, chenjj@tyut.edu.cn

Abstract. This paper proposes a frequent closed itemsets mining algorithm based on FP-tree and COFI-tree. This algorithm adopts a relatively small independent tree called COFI-tree. COFI-Tree doesn't need to construct conditional FP-Tree recursively and there is only one COFI-Tree in memory at a time, therefore this new mining algorithm reduces memory usage. Experiment shows that the new approach outperforms similar state-of-the-art algorithms when mining extremely large datasets in terms of execution time.

Keywords: Frequent Closed Itemsets, FP-Tree, COFI-Tree.

1 Introduction

Frequent pattern mining plays an important role in data mining. Now there are many mining algorithms about frequent closed itemsets, such as, CLOSET[1]. Through our study, we find this algorithm have some problems, it builds conditional FP-Tree recursively so that it requires more memory and CPU resources. In this paper, an effective frequent closed pattern mining algorithm is advanced, and it based on FP-Tree and COFI-Tree.

2 Related Work

There are many algorithms to address the problem of mining association rules [2], [6]. One of the important algorithms is the Apriori algorithm [6]. It also is the foundation of other most known algorithms. However, when mining extremely large datasets, the Apriori algorithm still suffers from two main problems of repeated I/O scanning and high computational cost. Another innovative approach of discovering frequent patterns, FP-Growth, was presented by Han et al. in [2]. It creates a compact tree-structure, FP-Tree, representing frequent patterns, that reduces the multi-scan times and improves the candidate itemset generation. The authors of FP-Tree algorithm have validated that their algorithm is faster than the Apriori. However, It needs to

construct conditional FP-Tree recursively. This massive creation of conditional trees makes FP-Tree algorithm not scalable to mine large datasets beyond few millions. The COFI-tree[3](Co-Occurrence Frequent Item Tree) algorithm that we are presenting in this paper is based on the core idea of the FP-Growth algorithm proposed by Han et al. in [2].

3 Construction of FP-Tree

A FP-Tree[2] is constructed by scanning database twice. First, a scan of database derives a list of frequent items, Second, the FP-Tree is constructed.

Definition 1 (FP-tree). A frequent-pattern tree is a tree structure

1. It consists of one root labeled as “null”, a set of item-prefix subtrees as the children of the root, and a frequent-item-header table.
2. Each node in the item-prefix subtree consists of three fields: item-name, count, and node-link, where item-name registers which item this node represents, count registers the number of transactions represented by the portion of the path reaching this node, and node-link links to the next node in the FP-tree carrying the same item-name, or null if there is none.
3. Each entry in the frequent-item-header table consists of two fields, (1) item-name and (2) head of node-link (a pointer pointing to the first node in the FP-tree carrying the item-name).

Algorithm1 (Construction of FP-tree)

Input: A transaction database DB and a minimum support threshold ξ .

Output: FP-tree.

Method: The FP-tree is constructed as follows.

1. Scan the transaction database DB once. Collect F, the set of frequent items, and the support of each frequent item. Sort F in support-descending order as L, the list of frequent items.
2. Create the root of an FP-tree, T, and label it as “null”. For each transaction in DB do the following.

Select the frequent items in transaction and sort them according to the order of L. Let the sorted frequent-item list in transaction be $[p \mid P]$, where p is the first element and P is the remaining list. Call insert-tree ($[p \mid P], T$).

The function insert tree ($[p \mid P], T$) is performed as follows.

If T has a child N such that $N.item-name = p.item-name$, then increment N’s count by 1; else create a new node N, with its count initialized to 1, its parent link linked to T, and its node-link linked to the nodes with the same item-name via the node-link structure. If P is nonempty, call insert-tree(P, N) recursively.

To understand effortlessly, let’s make an example.

Table 1. A transaction database (set the minimum support threshold as 3)

T1	T2	T3	T4
DE	ABDFEH	AGDECB	AGDFEI
T5	T6	T7	T8
AGEBFC	ADHEF	AGDBLC	ABCF
T9	T10	T11	
AADBCG	AFBCE	ABCH	

First, a scan of DB collects F , the set of frequent item and the support of each frequent item. $F = \{(A:10), (C:7), (D:7), (E:7), (F:6), (B:8), (G:5), (H:3)\}$, Sort F in support-descending order as L, the list of frequent items, $L = \{(A:10), (B:8), (C:7), (D:7), (E:7), (F:6), (G:5), (H:3)\}$.

Second, construction FP-Tree according to Algorithm 1, in the end, FP-Tree is constructed as Fig. 1.

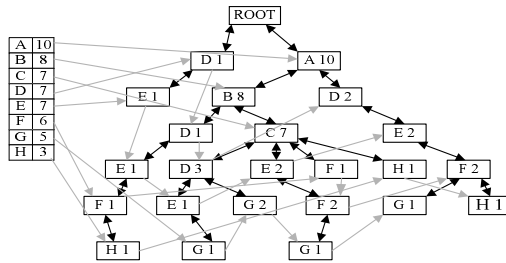


Fig. 1. FP-Tree

4 Construction of COFI-Tree

The COFI-trees are similar to the FP-Trees in general. However, the COFI-trees have bidirectional links in the tree allowing bottom-up scanning as well, and the nodes contain not only the item-name and a frequency counter, but also a participation counter. The COFI-tree for a given frequent item X contains only nodes labeled with items that are more frequent than or as frequent as X. The COFI-trees of all frequent items are not constructed together. Each tree is built, mined, and then discarded before the next COFI-tree is built. So at any given time, only one COFI-Tree is present in main memory.

Algorithm2 (construction of COFI-Tree)

Input: FP-Tree

Output: COFI-Tree

Method: The COFI-Tree is constructed as follows

1. Construct COFI-Tree for X. All branches can be found by following the chain of item X in the FP-Tree structure, and the support of each branch is equal to the support of the X node in its corresponding branch in FP-Tree.
2. The X-COFI-tree starts with the root node X. For each branch containing X do the following:

If multiple frequent items share the same prefix, they are merged into one branch and a counter for each node of the tree is adjusted accordingly, else a new branch is formed starting from the root node X and the support of this branch is equal to the support of the X node in its corresponding branch in FP-Tree. In each node, participation count is initialized to 0 and is used by the mining algorithm discussed later. Last two kinds of link are formed: a horizontal link which points to the next node that has the same item-name in the tree, and a bidirectional vertical link that links a child node with its parent and a parent with its child.

Fig.2. illustrates G- COFI-trees for frequent items of Fig.1. In Fig.2. the rectangle nodes are nodes from the tree with an item label and two counters. The first counter is a support-count for that node while the second counter, called participation-count, is initialized to 0.

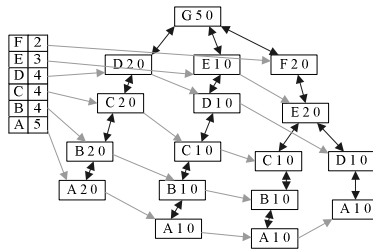


Fig. 2. G-COFI-Tree

4.1 Mining the COFI-Tree

Algorithm3 (Mining the COFI-Tree)

Input: COFI-Tree

Output: Candidate frequent patterns

Method: COFI-Tree is mined as follows:

For each frequent item in the header table of the COFI-Tree do as following:

Some branches that contain frequent item are found in the COFI-Tree, and the branch-support of each branch is the frequency of the first item in the branch minus the participation value of the same node. At the same time, the participation values for all nodes in this branch are incremented by the frequency of each branch.

For example, The G-COFI-tree in Fig.2. is mined for candidate frequent patterns. The process starts from the most globally frequent item, which is A, and then traverse all the A nodes if the support is greater than participation, the third counter on the node, then the complete path from this node to the COFI-root is built with branch-support equals to the difference between the support and participation of that node.

All values of participation for all nodes in these paths are updated with the participation of the original node . Candidate frequent patterns (A, B, C, D, G: 2), (A, B, C, E,G: 1), (A, D, E,G: 1), and (A, B, C, D, E,G: 1) are generated from this tree.

Steps to produce frequent patterns related to the G item for example, are illustrated in Fig.3.

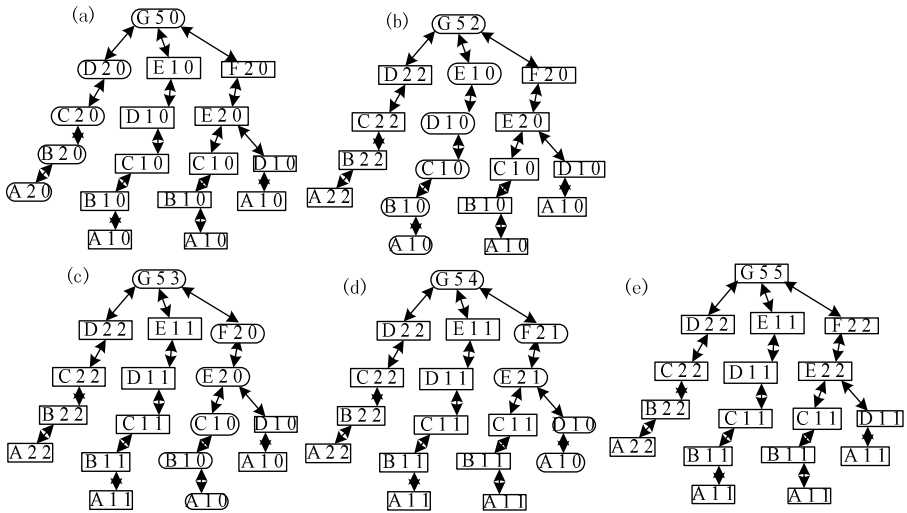


Fig. 3. The Process of finding candidate frequent patterns from G-COFI-Tree

5 Frequent Closed Itemsets

Definition 2 (Frequent closed itemsets)

An itemset Y is a frequent closed itemset if it is frequent and there exists no proper superset $Y' \supset Y$ such that $sup(Y') = sup(Y)$.

An array of pointer that has a size equal to the length of the largest candidate frequent patterns is adopted to derive frequent closed itemsets. Each node of the connected link list is made of 4 variables which are: the pattern, a pointer to the next node, and two number variables that represent the support and branch-support of this pattern. The support records the number of times this pattern occurs in the database. The branch-support records the number of times this pattern occurs alone without other frequent items.

Algorithm 4 (Frequent closed itemsets)

Input: Candidate frequent patterns

Output: Frequent closed itemsets

Method: Steps as following:

1. Candidate frequent patterns is stored into an array of pointer, and set support is equal to branch-support.
2. Intersect each one of these patterns with all other candidate frequent patterns to get a set of potential candidates, and potential candidates are stored into array of pointer, and support and branch-support are equal to 0.
3. Count the support of each generated patterns. The support of each one of them is the summation of supports of all its supersets of candidate frequent patterns.
4. Scan these patterns to remove non-frequent ones or frequent ones that already have a frequent superset with the same support.

As an example, candidate frequent patterns in section 3.1 are mined for frequent closed pattern. Candidate frequent patterns (A, B, C, D, G: 2), (A, B, C, E, G: 1), (A, D, E, G: 1), and (A, B, C, D, E, G: 1) are generated from G-COFI-Tree. According to algorithm 4, itemsets is changed into (ABCDEG:1,1), (ABCDG:2,2), (ABCEG:1,1), (ADEG:1,1) after step 1. Itemsets is changed into (ABCDEG:1,1), (ABCDG:2,2), (ABCDG:0,0), (ABCEG:1,1), (ABCEG:0,0), (ADEG:1,1), (ADEG:0,0), (ABCG:0,0), (ADG:0,0), (AEG:0,0) after step 2. Itemsets is changed into (ABCDEG:1,1), (ABCDG:2,2+1), (ABCDG:0,0+1+2), (ABCEG:1,1+1), (ABCEG:0,0+1+1), (ADEG:1,1+1), (ADEG:0,0+1+1), (ABCG:0,0+1+2+0+1+0), (ADG:0,0+1+2+0+1+0), (AEG:0,0+1+1+0+1+0) after step 3. In step 4, itemsets (ABCDEG:1,1) (ABCEG:1,2), (ABCEG:0,2) (ADEG:1,2) (ADEG:0,2), (ABCDG:0,3) are discarded. At last, frequent closed patterns (ABCDG), (ABCG), (ADG), (AEG) are generated and added to the pool of closed patterns. The G-COFI-tree and its array of pointers are cleared from memory as there is no need for them any more. The same process repeats with the remaining COFI-trees for F and E, where any newly discovered frequent closed pattern is added to the global pool of closed patterns.

5.1 Frequent Closed Pattern Mining Algorithm Based on COFI-Tree

Algorithm 5 (Frequent closed pattern mining algorithm based on COFI-Tree)

Input: A transactional database

Output: Frequent closed pattern

Method:

1. construction of FP-Tree
2. construction of COFI-Tree
3. mining COFI-Tree to find candidate frequent patterns
4. Travel candidate frequent patterns to derive frequent closed pattern.

6 Experiment

Some papers[4] had validated that FP-closed is better than CLOSET, so here a performance study is presented to evaluate the new approach against the state-of-art algorithm that mine closed patterns which is FP-Closed. The experiment is conducted on an HP Intel Core i5-2300 CPU 2.80GHz with 4GB memory. Experiment database is T10I4D100K[5].

6.1 Compare with FP-Closed

Experiment analyzing contains two aspects: First, analyze performance of two algorithms at spatial complexity by changing the support of transaction database. The comparing result is shown in Fig.4. Second, analyze performance of two algorithms at time complexity by changing the size of transaction database. The comparing result is shown in Fig.5.

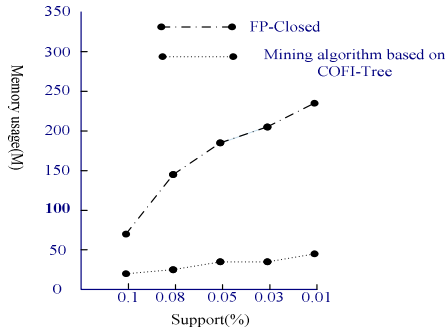


Fig. 4. Spatial complexity of two algorithms

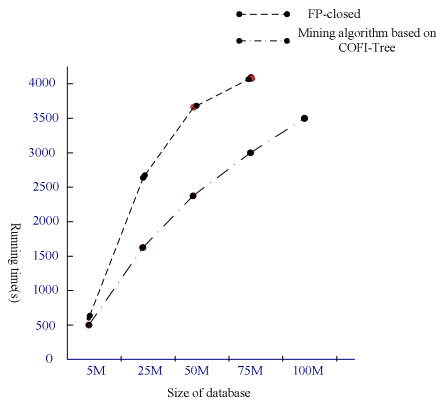


Fig. 5. Two complexity of two algorithms

6.2 Performance Evaluation

Fig.4. reveals the applicability and scalability of the COFI-tree algorithm. It shows that memory usages of FP-closed algorithm tends to grow exponentially when support threshold reduces from 0.1% to 0.01%. However, as support threshold reduces smaller and smaller, there is subtle change in memory usages of COFI-tree algorithm.

From Fig.5, it is easy to notice that COFI-tree algorithm has better performance for large data set. It is because this algorithm is no need to construct conditional sub-trees recursively. The experiments we conducted showed that our algorithm is scalable to mine tens of millions of transactions.

7 Conclusion

In this paper, a frequent closed pattern mining algorithm based on COFI-Tree is advanced. In this algorithm, COFI-Tree is adopted. COFI-Tree doesn't need to build conditional FP-Tree recursively and there is only one COFI-Tree in memory at a time, therefore this new mining algorithm reduces memory usage. The result of experiment shows that this algorithm is better than others on running time.

References

1. Pei, J., Han, J., Mao, R.: An efficient algorithm for mining frequent closed itemsets. In: Gunopulos, D., et al. (eds.) Proc. of the 2000 ACM SIGMOD Int'l. Workshop on Data Mining and Knowledge Discovery. ACM Press, Dallas (2000)
2. Han, J., Pei, J., Yin, Y.: Mining frequent patterns without candidate generation. In: 2000 ACM SIGMOD Intl. Conference on Management of Data, pp. 1–12 (2000)
3. El-Hajj, M., Zaiane, O.R.: Non recursive generation of frequent k-itemsets from frequent pattern tree representations. In: Proc. of 5th International Conference on Data Warehousing and Knowledge Discovery, DaWak 2003 (2003)
4. Leung, C.K.-S., Mateo, M.A.F., Brajczuk, D.A.: A Tree-Based Approach for Frequent Pattern Mining from Uncertain Data. In: Washio, T., Suzuki, E., Ting, K.M., Inokuchi, A. (eds.) PAKDD 2008. LNCS (LNAI), vol. 5012, pp. 653–661. Springer, Heidelberg (2008)
5. <http://fimi.cs.helsinki.fi/data/>
6. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules. In: Proc.1994 Int. Conf. Very Large Data Bases, Santiago, Chile, pp. 487–499 (1994)

Improved Algorithm for Mining N-Most Interesting Itemsets

Xiaohong Cui¹, Jihai Xiao¹, Junjie Chen², and Lijun Sang¹

¹ College of Textile Engineering and Art, Taiyuan University of Technology,
030600 JinZhong, China

² College of Computer Science and Technology, Taiyuan University of Technology,
030024 TaiYuan, China

{tycuixiaohong, xiaojihai}@126.com, chenjj@tyut.edu.cn,
slj9988@163.com

Abstract. BOMO algorithm constructs conditional FP-Tree recursively so that it requires more memory and CPU resources. To solve this problem, an algorithm for mining N-most interesting itemsets based on COFI-Tree is presented. This algorithm adopts COFI-Tree. COFI-Tree doesn't need to construct conditional FP-Tree recursively and there is only one COFI-Tree in memory at a time. Experiment shows that (1) the new algorithm based on COFI-tree performs faster than current best algorithm BOMO;(2) the algorithm has good performance for large data set, especially it shows the best when for k value is smaller than 4.

Keywords: Data mining, association rules, N-most interesting itemsets, FP-Tree, COFI-Tree.

1 Introduction

Agrawal and Srikant [1] first proposed an algorithm to tackle the problem of mining frequent itemsets with frequency greater than a given threshold. The bottleneck of the Apriori-like method is at the candidate set generation and test. A data structure called Frequent-Pattern tree, or FP-tree and an efficient algorithm called FP-growth are proposed by Han et al. [2] to overcome the above weaknesses. El-Hajj and Zaiane[3] found that FP-growth algorithm has a bottleneck of the excessive recursive call of functions. They thus proposed the COFI-tree which is similar to the FP-tree. El-Hajj and Zaiane [4] claim that the COFI-tree should have a better performance, when compared with the FP-tree. Cheung and Fu [5], [6] proposed an algorithm called BOMO which mines the N most interesting itemsets. Algorithm BOMO is based on an FP-tree structure. Fu et al. in [7] presents a novel technique of mining only N-most interesting frequent itemsets without specifying any minimum support. Han et al. in [8], proposed an another variation of mining Top-K frequent closed itemsets with length greater than a minimum user specified threshold , where K is a user-desired number of frequent closed itemsets to be mined.

Thus, in this paper, we propose an algorithm based on the COFI-tree in order to find N most interesting itemsets. The problem of mining N -most interesting frequent itemsets of size k at each level of $1 \leq k \leq k_{max}$, given N and k_{max} can be considered from the following definitions.

Table 1. Definitions

k_{max}	Upper bound on the size of interesting itemsets to be found
ξ	Current support threshold for all the itemsets
ξ_k	Current support threshold for the k -itemsets

Definition 1. *k-itemsets*: is a set of items containing k items.

Definition 2. The *N-most interesting k-itemsets*. Let us sort the k -itemsets by descending support values, let S be the support of the N -th k -itemset in the sorted list. The N -most interesting k -itemsets are the set of k -itemsets having support $\geq S$.

Definition 3. The *N-most interesting itemsets* is the union of the N -most interesting k -itemsets for each $1 \leq k \leq k_{max}$, where k_{max} is the upper bound of the size of itemsets we would like to find. We say that n itemset in the N -most interesting itemsets is interesting.

2 Method for Mining N-Most Interesting Itemsets by COFI-Tree

2.1 Construction of FP-Tree

A FP-Tree is constructed by scanning database twice. First, a scan of database derives a list of frequent items, Second, the FP-Tree is constructed.

Definition 4 (FP-tree)

A frequent-pattern tree is a tree structure.

1. It consists of one root labeled as “null”, a set of item-prefix subtrees as the children of the root, and a frequent-item-header table.
2. Each node in the item-prefix subtree consists of three fields: item-name, count, and node-link, where item-name registers which item this node represents, count registers the number of transactions represented by the portion of the path reaching this node, and node-link links to the next node in the FP-tree carrying the same item-name, or null if there is none.
3. Each entry in the frequent-item-header table consists of two fields, (1) item-name and (2) head of node-link (a pointer pointing to the first node in the FP-tree carrying the item-name).

Algorithm 1 Build (T,D, ξ)

Input: A transaction database D and a minimum support threshold ξ .

Output: FP-tree, T.

Method: The FP-tree is constructed as follows.

1. Scan the transaction database D once. Collect F, the set of frequent items, and the support of each frequent item. Sort F in support descending order as L, the list of frequent items.

2. Create the root of an FP-tree, T, and label it as “null”. For each transaction in D do the following.

Select the frequent items in transaction and sort them according to the order of L. Let the sorted frequent-item list in transaction be [p | P], where p is the first element and P is the remaining list. Call insert-tree ([p | P], T).

The function insert-tree ([p | P], T) is performed as follows. If T has a child N such that N.item-name = p.item-name, then increment N’s count by 1; else create a new node N, with its count initialized to 1, its parent link linked to T, and its node-link linked to the nodes with the same item-name via the node-link structure. If P is nonempty, call insert-tree (P, N) recursively.

2.2 Construction of COFI-Tree (Co-occurrence Frequent Item Tree)

The COFI-trees are similar to the FP-Trees in general. However, the COFI-trees have bidirectional links in the tree allowing bottom-up scanning as well, and the nodes contain not only the item-name and a frequency counter, but also a participation counter. The COFI-tree for a given frequent item X contains only nodes labeled with items that are more frequent than or as frequent as X. The COFI-trees of all frequent items are not constructed together. Each tree is built, mined, and then discarded before the next COFI-tree is built. So at any given time, only one COFI-Tree is present in main memory.

Algorithm 2 COFI (T, ξ)

Input: FP-Tree, a minimum support threshold ξ

Output: Full set of frequent patterns

Method:

1. B = the most frequent item on the header table of FP-Tree
2. While (There are still frequent items) do {
 - 2.1 counts the frequency of all items that share item (B) in a path. Frequencies of all items that share the same path are the same as of the frequency of the (B) items
 - 2.2 Remove all non-locally frequent items for the frequent list of item (B)
 - 2.3 Create a root node for the (B)-COFI-tree with both frequency-count and participation-count = 0

2.3.1 C is the path of locally frequent items in the path of item B to the root

2.3.2 Items on C form a prefix of the (B)-COFI-tree.

2.3.3 If the prefix is new then

Set frequency-count=frequency of (B) node and participation-count=0 for all nodes in the path

Else

2.3.4 Adjust the frequency-count of the already existing part of the path.

2.3.5 Adjust the pointers of the Header list if needed

2.3.6 Find the next node for item B in the FP-tree and go to 2.3.1

2.4 Mine (B)-COFI-tree

2.5 Release (B) COFI-tree

2.6 B = next frequent item from the header table}.

2.3 Mining COFI-Tree

Algorithm 3 Mine (B)-COFI-tree

Input: B-COFI-tree

Output: part of most interesting itemsets

Method:

1 node B = select_next_node

(Selection of nodes starting with the node of the most locally frequent item and following its chain, until we reach the least frequent item in the Header list of the (B)-COFI-tree)

2 While there are still nodes do {

2.1 D = set of nodes from node B to the root

2.2 F = nodeB.frequency-count – nodeB.participation-count

2.3 Generate all Candidate patterns X from items in D. Patterns that do not have B will be discarded.

2.4 Patterns in X that do not exist in the B-Candidate List will be added to it with frequency = F otherwise just increment their frequency with F

2.5 Increment the value of participation-count by F for all items in D

2.6 node B = select_next_node}

3 Based on support threshold ξ remove non-frequent patterns from B-Candidate List.

4 For each pattern S in the B-Candidate List

4.1 L = number of items in pattern S

4.2 If ($L < k_{\max}$) {

4.2.1 If frequency of pattern S $\geq \xi_L$ then

4.2.1.1 Insert pattern S into resultL; update ξ_L ; update ξ if necessary.

{ $\xi = \min(\xi_1, \xi_2, \dots, \xi_{k_{\max}})$ }

4.2.2 If ($\xi_L \neq 0$) and (resultL contains itemset X with support < frequency of pattern S),

4.2.3 Remove X from resultL}.

2.4 Pseudo Code of Mining N-Most Interesting Itemsets

Pseudo Code of mining N-most interesting itemsets is following:

1. Let $result_k$ be the resulting set of interesting k -itemsets. Set $result_k = \Phi$
2. $\xi_1 = \xi_2 = \xi_3 = \xi_4 = \dots = \xi_{k_{max}} = \xi = 0$
3. Create FP-tree, T, with root node only labeled as 'NULL'
4. Build (T, D, ξ)
5. COFI (T, $\xi, \xi_1, \xi_2, \dots, \xi_{k_{max}}$).

3 Experiment

The experiment is conducted on a HP Intel Core i5-2300 CPU 2.80GHz with 4GB memory. It used the benchmark datasets available at <http://fimi.ua.ac.be/data/>. These datasets are frequently used in many frequent itemsets mining algorithms. We label our dataset by Dx.Ty.Mz, where D refers to the Number of transactions, T is the average number of items in a transaction, M is the number of different items, for dataset Dx.Ty.Mz, there are x transactions, each transaction has y items on average and there are z different possible items.

Experiment analyzing contains two aspects: First, analyze performance of the new algorithm based on COFI-Tree by changing the size of transaction database. The result is shown in Fig.1 and 2. Second, analyze performance of two algorithms at time complexity by comparing with BOMO on same condition. The comparing result is shown in Fig.3.

3.1 Performance of New Algorithm

Fig. 1. Fix $k = (4, 6)$, varying $N = (10, 20, 40, 100)$ using D100k.T10.M1k.

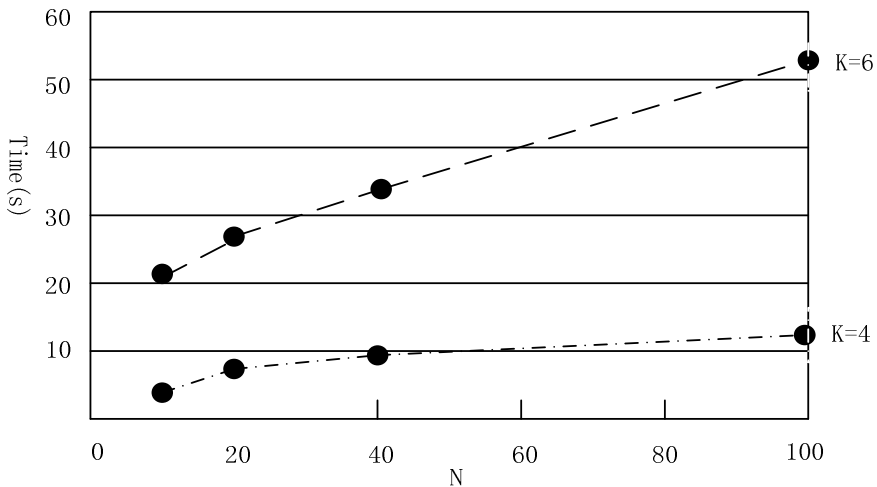


Fig. 1. Performance results on D100k.T10.M1k

Fig.1 show that for a fix k, the total time needed to mine a result set grows almost linearly with N.

Fig. 2. Fix N = (20, 100), various k = (1, 2, 3, 4, 5, 6, 7), using D1k.T10.M1k

Fig.2 reveals that for $k < 4$, the mining time grows linearly for a wide range of N. However, as k grows larger and larger, especially when $k > 6$, the mining time tends to grow exponentially.

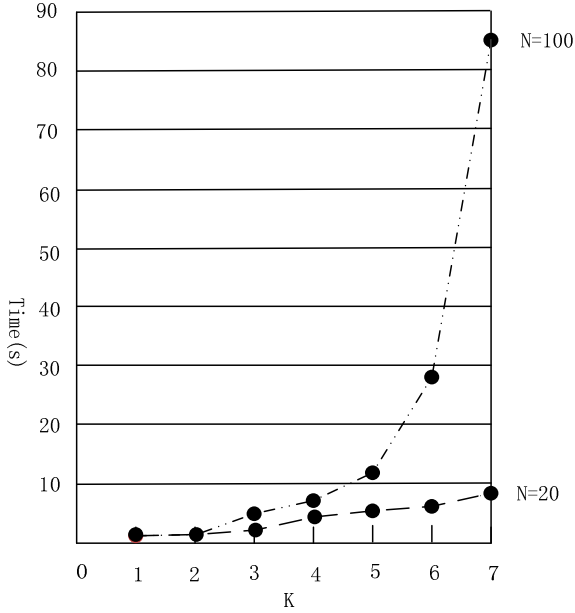


Fig. 2. Performance results on D1k.T10.M1k

3.2 Compare with BOMO

Fig. 3. Fix k = (4, 6), various N = (10, 20, 40, 100) using D100k.T10.M1k.

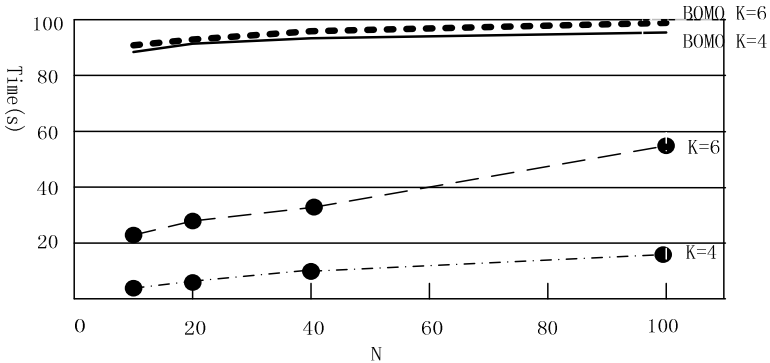


Fig. 3. Compare with BOMO

From Fig.3, it is easy to notice that this algorithm has better performance for all test cases for large data set. It is because this algorithm has a good pruning power and there is no need to construct conditional sub-trees recursively.

4 Conclusion

We have implemented an algorithm to solve the problem of mining N most interesting k-itemsets without giving a minimum support threshold. This algorithm based on COFI-tree performs faster than current best algorithms BOMO, especially well with small k.

References

1. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules. In: Proceedings of the 20th VLDB Conference, pp. 487–499 (1994)
2. Han, J., Pei, J., Yin, Y.: Mining frequent patterns without candidate generation. In: 2000 ACM SIGMOD Intl. Conference on Management of Data, pp. 1–12 (2000)
3. El-Hajj, M., Zaïane, O.R.: COFI-tree mining: a new approach to pattern growth with reduced candidacy generation. In: Workshop on Frequent Itemset Mining Implementations (FIMI 2003) in Conjunction with IEEE-ICDM (2003)
4. El-Hajj, M., Zaiane, O.R.: Non recursive generation of frequent k-itemsets from frequent pattern tree representations. In: Proceeding of 5th International Conference on Data Warehousing and Knowledge Discovery, DaWak 2003 (2003)
5. Cheung, Y.L., Fu, A.W.: An FP-tree Approach for Mining N-most Interesting Itemsets. In: Proceedings of the SPIE Conference on Data Mining (2002)
6. Cheung, Y.L., Fu, A.W.-C.: Mining association rules without support threshold: with and without item constraints. *IEEE Transactions on Knowledge and Data Engineering (TKDE)* 16, 1052–1069 (2004)
7. Fu, A.W.-c., Kwong, R.W.-w., Tang, J.: Mining N-most Interesting Itemsets. In: Ohsuga, S., Raś, Z.W. (eds.) ISMIS 2000. LNCS (LNAI), vol.1932, pp. 59–67. Springer, Heidelberg (2000)
8. Wang, J., Han, J., Lu, Y., Tzvetkov, P.: TFP: An Efficient Algorithm for Mining Top-K Frequent Closed Itemsets. *IEEE Transaction on Knowledge and Data Engineering* 17(5), 652–664 (2005)

An Improved KFCM Algorithm Based on Artificial Bee Colony

Xiaoqiang Zhao* and Shouming Zhang

College of Electrical and Information Engineering, Lanzhou University of Technology,
Lanzhou, 730050, China

Key Laboratory of Gansu Advanced Control for Industrial Processes, Lanzhou,
730050, China

xqzhao@lut.cn, zhsm-2008@163.com

Abstract. Kernel fuzzy C-means (KFCM) clustering Algorithm is one of the most widely used methods in data mining, but this algorithm still exists some defects, such as the local optima and sensitivity to initialization and noise data. Artificial bee colony (ABC) is a very simple, robust, stochastic global optimization tool which is used in many optimization problems. In this paper, an improved KFCM algorithm based on ABC (ABC-KFCM) is proposed. It can integrate advantages of KFCM and ABC algorithm. According to the test, compared with the FCM and KFCM clustering algorithm, the proposed algorithm improves the optimization ability of the algorithm, the number of iterations is fewer, and the convergence speed is faster. In addition, there is also a large improved in the clustering result.

Keywords: Data mining, kernel fuzzy C-mean clustering, artificial bee colony, ABC-KFCM.

1 Introduction

With development of era, the number of data in various ways has increased rapidly. It is difficult to take full advantage of the useful knowledge stored in these data by the traditional approaches. Data mining technology emerges as the times require and is abroad applied on every field. Data Mining is defined as a process that gets information and knowledge which are connotative, unknown and useful from practical data which is substantial, incomplete, noise, ambiguous and stochastic[1].

During the past decade, clustering analysis as one of the main method of data mining causes the attention of people more and more. Clustering is the process of assigning data objects into a set of disjoint groups called clusters so that objects in each cluster are more similar to each other than objects from different clusters.

* Corresponding author: Xiaoqiang Zhao. This works is supported by Natural Science Foundation of Gansu Province (No. 0809RJZA005) and Foundation of Advanced Control Key Laboratory of Industrial Process of Gansu Province (No.XJK0907).

K-means [2] is one of the most popular hard clustering algorithms which partitions data objects into k clusters where the number of clusters. This model is inappropriate for real data sets in which there are no definite boundaries between the clusters.

Since Zadeh proposed fuzzy sets that introduced the idea of partial memberships described by membership functions, it has been successfully applied in various areas. Especially, fuzzy sets could allow membership functions to all clusters in a data set so that it was very suitable for cluster analysis[3]. Ruspini first proposed fuzzy c -partitions as a fuzzy approach to clustering. Later, the fuzzy c -means (FCM) algorithms with a weighting exponent $m=2$ proposed by Dunn, and then generalized by Bezdek with $m>1$ became popular[4]. Fuzzy c -means clustering is an effective algorithm, but the random selection in center points makes iterative process falling into the local optimal solution easily.

Recently, tremendous works focus on using kernel method, which first maps the data into high dimension space to gain high discriminant capability, and then calculates the measure of the samples in their original data space with kernel function. Kernel fuzzy C-Means (KFCM) is proposed by substituting the Euclidean distance with kernel function.

KFCM not only to certain extent overcomes limitation of data intrinsic shape dependence and can correctly clustering, but also overcome sensitivity to initialization and noise data and improve the algorithm robustness. However, like FCM algorithm, KFCM algorithm still exists some drawbacks, such as the sensitivity to initialization and the tendency to get trapped in local minima. Therefore, an improved kernel fuzzy C-Means based on artificial bee colony (ABC-KFCM) is put forward.

2 KFCM Algorithm

FCM partitions a given dataset $X = \{x_1, x_2, \dots, x_n\} \in R^p$, into c fuzzy subsets by minimizing the following objective function

$$J_m(U, V) = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m \|x_k - v_i\|^2 \tag{1}$$

where c is the number of clusters and selected as a specified value, n the number of data points, u_{ik} the membership of x_k in class i , m the quantity controlling clustering fuzziness, and V the set of cluster centers ($v_i \in R^p$). The matrix U satisfies

$$U \in \left\{ u_{ik} \in [0,1] \mid \sum_{i=1}^c u_{ik} = 1, \forall k \quad \text{and} \quad 0 < \sum_{k=1}^n u_{ik} < N, \forall i \right\} \tag{2}$$

Define a nonlinear map as $\Phi: x \rightarrow \Phi(x) \in F$, where $x \in X$. X denotes the data space, and F the transformed feature space with higher or even infinite dimension. KFCM minimizes the following objective function

$$J_m(U, V) = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m \|\Phi(x_k) - \Phi(v_i)\|^2 \tag{3}$$

Where $\|\Phi(x_k) - \Phi(v_i)\|^2 = K(x_k, x_k) + K(v_i, v_i) + 2K(x_k, v_i)$ (4)

Where $K(x, y) = \Phi(x)^T \Phi(y)$ is an inner product kernel function. If we adopt the Gaussian function as a kernel function, i.e., $K(x, y) = \exp(-\|x - y\|^2 / \sigma^2)$, then $K(x, x) = 1$, according to Eq.(3) and Eq.(4), can be simplified to

$$J_m(U, V) = 2 \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m (1 - K(x_k, v_i)) \tag{5}$$

Minimizing Eq.(5) under the constraint of U , we have

$$u_{ik} = \frac{(1/(1 - K(x_k, v_i)))^{1/(m-1)}}{\sum_{j=1}^c (1/(1 - K(x_k, v_j)))^{1/(m-1)}}, \forall i = 1, 2, \dots, c; k = 1, 2, \dots, n \tag{6}$$

$$V_i = \frac{\sum_{k=1}^n u_{ik}^m K(x_k, v_i) x_k}{\sum_{k=1}^n u_{ik}^m K(x_k, v_i)}, \forall i = 1, 2, \dots, c \tag{7}$$

The proposed kernelized fuzzy C-means algorithm can be summarized in the following steps[5]:

Step 1: Set c, t_{max}, m and $\epsilon > 0$ for some positive constant.

Step 2: Initialize the membership matrix u_{ik}^0 .

Step 3: For $t = 1, 2, \dots, t_{max}$, do:

update the cluster centers V_i^t with Eq. (7);

update membership matrix u_{ik}^t with Eq. (6);

compute $E^t = \max_{i,k} |u_{ik}^t - u_{ik}^{t-1}|$, if $E^t \leq \epsilon$, stop;

end.

3 Artificial Bee Colony Algorithm

Artificial Bee Colony (ABC) algorithm was proposed by Karaboga for optimizing numerical problems in [6]. The algorithm simulates the intelligent foraging behavior of honey bee swarms. It is a very simple, robust and population based stochastic optimization algorithm.

In ABC[7] algorithm, the colony of artificial bees consists of three groups of bees: employed bees, onlookers and scouts. A food source represents a possible solution to the problem to be optimized and the nectar amount of a food source corresponds to the quality (fitness) of the associated solution, calculated by:

$$fit_i = \frac{1}{1 + f_i} \tag{8}$$

In the algorithm[7][8], the number of the employed bees or the onlooker bees is equal to the number of solutions (the cluster centers) in the population. At the first step, the ABC generates a randomly distributed initial population of SN (the size of population) solutions (food source positions), Each solution x_i ($i=1, 2, \dots, SN$) is a D(the number of optimization parameters)dimensional vector. After initialization, the population of the positions is subjected to repeated cycles, $C = 1, 2, \dots, MCN$, of the search processes of the employed bees, the onlooker bees and scout bees. Each employed bee moves onto her food source area for determining a new food source within the neighborhood of the present one, and then evaluates the nectar amount. If the nectar amount of the new one is higher, the bee memorizes the new position and forgets the old one. After all employed bees complete the search process, they share the nectar information of the food sources and their position information with the onlooker bees on the dance area. An onlooker bee evaluates the nectar information taken from all employed bees and chooses a food source with a probability related to its nectar amount. If the nectar amount of the new one is higher, the bee memorizes the new position and forgets the old one.

An artificial onlooker bee chooses a food source depending on the probability value associated with that food source, p_i , calculated by:

$$p_i = \frac{fit_i}{\sum_{n=1}^{SN} fit_n} \tag{9}$$

In order to produce a candidate food position from the old one in memory, the ABC uses the following expression:

$$v_{ij} = x_{ij} + \phi_{ij}(x_{ij} - x_{kj}) \tag{10}$$

Where $j \in \{1, 2, \dots, D\}$ and $k \in \{1, 2, \dots, SN\}$ are randomly chosen indexes.

Although k is determined randomly, it has to be different from i . ϕ_{ij} is a random number between [-1, 1]. As can be seen from (10), as the difference between the parameters of the x_{ij} and x_{kj} decreases, the perturbation on the position x_{ij} decreases, too. Thus, as the search approaches to the optimum solution in the search space, the step length is adaptively reduced.

In ABC, providing that a position can not be improved further through a predetermined number of cycles (called limit), then that food source is assumed to be abandoned. Assume that the abandoned source is x_i , then the scout discovers a new food source to be replaced with x_i . This operation can be defined as in

$$x_i^j = x_{min}^j + rand(0,1)(x_{max}^j - x_{min}^j) \tag{11}$$

ABC algorithm is a robust search process, exploration and exploitation processes are carried out together. The global search performance of the algorithm depends on random search process performed by scouts and neighbor solution production mechanism performed by employed and onlooker bees. Therefore, ABC algorithm is

an efficient optimization tool since it combines exploitative local search and explorative global search processes efficiently.

4 ABC-KFCM Algorithm

Comparison of clustering and bee colony foraging is shown as Table1.

Table 1. Comparison of clustering and bee colony foraging

colony foraging	clustering
Position of a food source	A possible solution(clustering center)
The nectar amount of a food source	The quality of the associated solution
Speed of bee forage	Solution speed
The maximizing rewards	The best clustering result

Let $X = \{x_1, x_2, \dots, x_n\}$ be a set of n objects, where x_i is a D-dimensional vector. In ABC, a bee denotes a cluster center, i.e. $V = \{v_1, v_2, \dots, v_c\}$, where v_j is also a D-dimensional vector. In ABC-KFCM algorithm the same as other evolutionary algorithms, we need a function for evaluating the generalized solutions called fitness function. In this paper, Eq.(12) is used for evaluating the solutions.

$$fit_i = \frac{1}{1 + J_m(U, V)} \tag{12}$$

Where $J_m(U, V)$ is objective function of KFCM algorithm given in Eq.(5). The smaller is $J_m(U, V)$, the higher is the individual fitness fit_i and the better is the clustering result.

ABC-KFCM algorithm uses the capacity of global search in ABC algorithm to seek optimal solution as initial clustering-centers of KFCM algorithm, and then use KFCM algorithm to optimize initial clustering-centers, so as to get the global optimum. Detailed description of each step is given below:

Step 1: Initialize the parameters of ABC and KFCM including population size SN, maximum cycle number MCN, limit, clustering number c, m and ϵ ;

Step 2: Compute kernel matrix $K(x_k, v_i)$ and initialize the membership matrix U^0 by Eq. (6)

Step 3: Generate the initial population (cluster center) c_{ij} by Eq. (7), and evaluate the fitness of the population by Eq. (12)

Step 4: ABC algorithm

4.1 Set cycle to 1

4.2 Set s to 1

4.3 FOR each employed bee {
 Produce new solution v_{ij} by using (6)
 Calculate the value fit_i
 Apply greedy selection process}

4.4 Calculate the probability values p_i for the solutions (c_{ij}) by (9)

4.5 FOR each onlooker bee {
 Select a solution c_{ij} depending on p_i
 Produce new solution v_{ij}
 Calculate the value fit_i
 Apply greedy selection process}

4.6 If the searching times surrounding an employed bee s exceeds a certain threshold limit, but still could not find better solutions, then the location vector can be reinitialized randomly according to Eq. (11), go to step 4.2

4.7 If the iteration value is larger than the maximum number of the iteration (that is, $cycle > MCN$), output the best cluster centers. If not, go to Step 4.1.

Step 5: KFCM algorithm

5.1 Update membership matrix u_{ik}^t with Eq. (6);

5.2 Update the cluster centers V_i^t with Eq. (7);

5.3 Compute $E^t = \max_{i,k} |u_{ik}^t - u_{ik}^{t-1}|$, if $E^t \leq \epsilon$, stop; If not, go to Step 5.1.

5 Experiment Results

Three classification problems from UCI database which is a well-known database repository, are used to evaluate the ABC clustering algorithm. The three datasets are well-known iris, wine and glass datasets taken from Machine Learning Laboratory. The experimental data sets can be found in Table 2.

Table 2. Experimental data sets

Data set name	Number of sample	Class	Dimension
IRIS	150	3	4
wine	178	3	13
glass	214	6	9

Iris dataset is perhaps the best-known database to be found in the pattern recognition literature. The data set contains three categories of 50 objects each, where each category refers to a type of iris plant. One category is linearly separable from the other two; the latter are not linearly separable from each other. There are 150 instances with four numeric features in iris data set. There is no missing attribute value. The attributes of the iris data set are sepal length, sepal width, petal length and petal width.

Wine dataset contains chemical analysis of 178 wines, derived from three different cultivars. Wine type is based on 13 continuous features derived from chemical analysis: alcohol, malic acid, ash, alkalinity of ash, magnesium, total phenols, flavanoids, nonflavanoid phenols, proanthocyaninism, color intensity, hue, OD280/OD315 of diluted wines and praline. The quantities of objects in the three categories of the data set are 59, 71 and 48, respectively.

Glass dataset is another biggest number of classes (6 classes) in the problems that we tackle. It is used to classify glass types as float processed building windows, non-float processed building windows, vehicle windows, containers, tableware, or head lamps. Nine inputs are based on 9 chemical measurements with one of 6 types of glass which are continuous with 70, 76, 17, 13, 9, and 29 instances of each class, respectively.

There are three control parameters in ABC algorithm, the swarm size SN, the maximum cycle number MCN and the limit. They are set as follow: SN=20, MCN=2000, limit=100. The weighting exponent m is set to 2.

These data sets cover examples of data of low, medium and high dimensions. For every dataset, algorithms performed 20 times individually for their own effectiveness tests, each time with randomly generated initial solutions. The experimental results are shown in table 3, table 4 and table 5.

As shown in these tables, the hybrid ABC-KFCM obtained superior results than others in all of data sets. ABC-KFCM's search ability has being enhanced, and its optimize speed is faster. Also the experimental results show that when the size of data set (number of objects or clusters) is large, the new algorithm has the better clustering results.

Table 3. Clustering result of IRIS data

Algorithms	Average Error Number			Average Accuracy Rate (%)	Iterations
	Setosa	Versicolo	Virginica		
ABC-KFCM	0/50	2/50	2/50	97.33	10
KFCM	0/50	5/50	6/50	92.67	13
FCM	0/50	6/50	10/50	89.33	15

Table 4. Clustering result of wine data

Algorithms	Average Error Number			Average Accuracy Rate (%)	Iterations
	class1	class2	class3		
ABC-KFCM	3/59	8/71	2/48	92.70	12
KFCM	7/59	12/71	3/48	87.64	15
FCM	10/59	16/71	5/48	82.58	22

Table 5. Clustering result of glass data

Algorithms	Average Error Number						Average Accuracy Rate (%)	Iterations
	1	2	3	4	5	6		
ABC-KFCM	3/70	3/76	117	1/13	0/9	2/29	95.33	10
KFCM	6/70	8/76	3/17	1/13	1/9	3/29	89.72	15
FCM	9/70	10/76	4/17	2/13	2/9	5/29	85.05	17

6 Conclusions

Artificial bee colony algorithm is a new, simple and robust optimization technique. In this paper, an ABC algorithm is developed to solve clustering problems which is inspired by the bees' forage behavior. ABC-KFCM algorithm uses ABC algorithm to seek optimal solution as initial clustering-centers of KFCM algorithm, and then use KFCM algorithm to optimize initial clustering-centers, so as to get the global optimum. Above all, it solves the problems of KFCM. Experimental results show that the new algorithm is more accurate in clustering, higher efficiency and fewer the number of iterations.

References

1. Fayyad, U.M., Piatsky-Shapiro, G., Smyth, P., Uthurusamy, R.: *Advances in Knowledge Discovery and Data Mining*, 1st edn. AAAI Press, Menlo Park (1996)
2. Smith, T.F., Waterman, M.S.: Identification of Common Molecular Subsequences. *J. Mol. Biol.* 147, 195–197 (1981)
3. Izakian, H., Abraham, A.: Fuzzy clustering using hybrid c-means and fuzzy particle swarm optimization. In: *Proceedings of the 2009 World Congress on Nature and Biologically Inspired Computing*, Coimbatore, pp. 1690–1694 (2009)
4. May, P., Ehrlich, H.C., Steinke, T.: ZIB Structure Prediction Pipeline: Composing a Complex Biological Workflow through Web Services. In: Nagel, W.E., Walter, W.V., Lehner, W. (eds.) *Euro-Par 2006. LNCS*, vol. 4128, pp. 1148–1158. Springer, Heidelberg (2006)
5. Yang, M.S., Tsai, H.S.: A gaussian kernel-based fuzzy c-means algorithm with a spatial bias correction. *Pattern Recognition Lett.* 29, 1713–1725 (2008)
6. Foster, I., Kesselman, C.: *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann, San Francisco (1999)
7. Liu, J., Xu, M.: Kernelized fuzzy attribute C-means clustering algorithm. *Fuzzy Sets Syst.* 159, 2428–2445 (2008)
8. Czajkowski, K., Fitzgerald, S., Foster, I., Kesselman, C.: Grid Information Services for Distributed Resource Sharing. In: *10th IEEE International Symposium on High Performance Distributed Computing*, pp. 181–184. IEEE Press, New York (2001)
9. Zhang, D.Q., Chen, S.C.: A novel kernelized fuzzy C-means algorithm with application in medical image segmentation. *Artificial Intelligence Med.* 32, 37–50 (2004)

10. 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)
11. Karaboga, D.: An idea based on honey bee swarm for numerical optimization. Technical report-TR06, Erciyes University, Engineering Faculty, Computer Engineering Department, Kayseri/Turkiye (2005)
12. Karaboga, D., Ozturk, C.: A novel clustering approach: Artificial bee colony (ABC) algorithm. *Applied Soft Computing Journal* (2008), doi:10.1016/j.asoc.2009.12.025
13. Zhang, C., Ouyang, D., Ning, J.: An artificial bee colony approach for clustering. *Expert Systems with Applications* 37, 4761–4767 (2010)

Improved Clustering Algorithm Based on Local Agglomerative Characteristics

Xi-xian Niu¹ and Yan-ping Cui²

¹ Faculty of Information Technology and Propagation, Hebei Youth Administrative Cadres College, Shijiazhuang, China

² College of Mechanical & Electronic Engineering, Hebei University of Science and Technology, Shijiazhuang, China

niuxixian@126.com, cuiyypkd@163.com

Abstract. Similarity measurement is the bases of clustering analysis. Current clustering algorithms mostly based on density and distance measurement, but these concepts become increasingly difficult to fit more and more complex data set and analysis works. SNN similarity, however, show more flexible ability to deal with different density, shape and multi-dimensions data process problems .In this paper we review mostly popular SNN based clustering method, give the definition of Local Agglomerative Characteristics during the procedure of the clustering, proposed a new clustering algorithm, that is, Improved Clustering algorithm based on Local Agglomerative Characteristics. Apply this clustering algorithm on experimental data set, the result show that it can work well on different type's data objects, can find nature distribute clusters in target data set, can improve the quality of data clustering.

Keywords: Data mining, clustering, SNN density, SNN similarity, local agglomerative characteristics.

1 Introduction

Due to rapid technological development in such areas as computer, network and communication, the researchers face ever-increasing challenges in extracting relevant information from the enormous volumes of available data. The so-called data avalanche is created by the fact that there is no concise set of parameters that can fully describe a state of real-world complex systems studied nowadays by biologists, ecologists, sociologists, economists, etc [1]. Pattern recognition is a primary conceptual activity of the Human being. Even without our awareness, clustering on the information that is conveyed to us is constant [2]. Clustering research has long been a hot research field of Data Mining, is considered the most important unsupervised learning problems. Specifically, clustering techniques are almost indispensable as a tool for data mining [2]. Clustering delineates operation for objects within a dataset having similar qualities into homogeneous groups, it allows for the discovery of similarities and differences among patterns in order to derive useful conclusions about them [3]. Purpose of the Clustering study is to determine the data groups in unlabeled data collection based on its inherent nature.

Cluster analysis is a challenging task and there are a number of well-known issues associated with it, e.g., finding clusters in data where there are clusters of different shapes, sizes, and density or where the data has lots of noise and outliers [1]. Determining the structure or patterns within data set is a significant works. Clustering depends critically on density and distance (similarity), but these concepts become increasingly more difficult to define as destination of clustering jobs become more complexity. As a consequence, find new measurement method is a very necessary task for clustering researchers. Hence, SNN(Shared Nearest Neighbors) base similarity measure is proposed, and several more efficient clustering algorithm are designed and implemented, such as, Jarvis-Patrick (SNN similarity based), SNN density based DBSCAN (Density-Based Spatial Clustering of Application with Noise). But these methods still have space to improved, so we can take local cluster feature into account, define new measurement, get more efficient clustering method. Therefore, the purpose of this paper is to analysis existing algorithms' limitation, find its resolution, and design improved algorithm.

2 SNN Similarity Based Jarvis-Patrick Clustering Algorithm

In the Jarvis-Patrick (JP) scheme, a shared nearest neighbor graph is constructed from the proximity matrix as follows. A link is created between a pair of point p and q if and only if p and q have each other in their closest k nearest neighbor lists. This process is called k -nearest neighbor sparsification. The weights of the links between two points in the SNN graph can either be simply the number of near neighbors which shared by the two points, or one can use a weighted version that takes the ordering of the near neighbors into account.

The JP clustering algorithm replaces the proximity between two points with the SNN similarity, which is calculated as described in algorithm 2. A threshold is then used to sparsify this matrix of similarities. At this point, all edges with weights less than a user specified threshold are removed and all the connected components in the resulting graph are our final clusters.

Algorithm 2: JP clustering algorithm

Compute the SNN similarity graph.

Sparsify the SNN similarity graph by applying a similarity threshold.

Find the connected components (clusters) of the sparsified SNN similarity graph.

Because JP clustering is based on the notion of SNN similarity, it is good at dealing with noise and outliers and can handle clusters of different sizes, shapes, and densities. The algorithm works well for high-dimensional data and is particularly good at finding tight clusters of strongly related objects.

A major drawback of the Jarvis – Patrick scheme is that, the threshold needs to be set high enough since two distinct set of points can be merged into same cluster even if there is only one link across them. On the other hand, if the threshold is too high, then a natural cluster may be split into too many small clusters due to natural variations in the similarity within the cluster. Another potential limitation is that not all objects are clustered [4] [6].

3 SNN Density Based Clustering Analysis

The SNN density method can be combined with DBSCAN algorithm to create a new clustering algorithm. This algorithm is similar to the JP clustering algorithm in that it starts with the SNN similarity graph. However, instead of using a threshold to sparsify the SNN similarity graph and then take connected components as clusters. The SNN algorithm, as DBSCAN, is a density-based clustering algorithm. The main difference between this algorithm and DBSCAN is that it defines the similarity between points by looking at the number of nearest neighbors that two points share. Using this similarity measure in the SNN algorithm, the density is defined as the sum of the similarities of the nearest neighbors of a point. Points with high density become core points, while points with low density represent noise points. All remainder points that are strongly similar to a specific core points will represent a new clusters.

The SNN density clustering needs three inputs parameters: K, the neighbors' list size; Eps, the threshold density; MinPts, the threshold that define the core points. After defining the input parameters, the SNN algorithm first finds the K nearest neighbors of each point of the dataset; second apply DBSCAN with user-specified parameters for Eps and MinPts to determine what the data objects' point type should be, that is, core, border, or noise. Two different approaches can be used to implement the SNN algorithms. One that creates the clusters around the core points previously identified. In the other approach the clusters are identified by the points that are connected in a graph. The graph is constructed by linking all the points which similarity is higher than Eps.

The algorithm automatically determined the number of clusters in data space, but not all the points are clustered. SNN density-based clustering finds clusters in which the points strongly related to one another. The strengths and limitations of SNN density-based clustering are similar to those of JP clustering. However, the use of core points and SNN density adds considerable power and flexibility to this approach [6] [7].

4 Improved Clustering Algorithm Based on Local Agglomerative Characteristics

SNN similarity measurement and SNN density based measurement, are both based on the local data space features, and take local configuration characteristics into account, mainly focus on the algorithm's adaptability to the problems, such as the overall context of density, different shapes and sizes etc. Improved LAC (local agglomerative characteristics) clustering algorithm, however, mainly focus on local clusters' configuration and distribution characteristics which displaying in the procedure of the whole clustering analysis, and its application. Analyze the features of the shared neighbors around the data object, such as density, size, shape etc, and redefine data objects' similarity measure, and then improve the algorithm's adaptability and optimize efficiency. The shared nearest neighbors between data objects is a local relatively tight cluster contrast to other data objects that scattered in the local area, since study on its data deploy characteristics is an very meaningful work to determine the data objects' similarity and density.

4.1 Similarity Analysis Based on LAC

Due to JP algorithm take rigid parameter k as the threshold to adjust data objects similarity computation, then it is disadvantage to find local area relative small cluster which has more tightly feature, these situation shown in figure 1(a), data object A and B will be grouped into the same cluster, even though their local view is relatively loosely like, while the parameter k is set equal to 6. In addition, because SNN density based algorithm take strict Eps as input parameter, then, it can't deal with some conditions, it's shown in figure 1(b), here, the objects A and B in the local configuration is more loosely while view directly, but within the limits in the Eps, but data objects C, compared to the object A, it has better connectivity to object B, although this connection is not direct.

In order to better reflect the local distribution of data collection how to affect the data objects' similarity computation; we can assess the data objects' similarity by the following aspects, such as relative density, local distribute shape, and local distance etc. in this notion, we can get the data objects have higher similarity in local area if the shared nearest neighbors have a relatively high density, or if the assessed data objects have a relatively short distance according to the distribute shape of their shared neighbors.

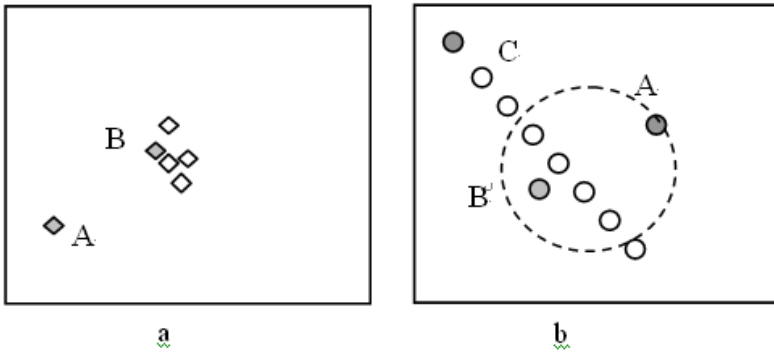


Fig. 1. Data objects' local configuration analysis

4.2 LAC's Definition and Its Measurement

Compared to other objects in target data collection, we can view the whole shared nearest neighbors as a cluster which has relative strong concentration properties in local data space. Therefore, the local clustering properties which get from the shared nearest neighbors can be seen as a measure basis to determine the assessed two objects whether or not have high similarity. In local data area, we can simple consider the two data objects have high similarity, if they have a relatively close distance. Because the distribution of data objects maybe exit different ways, in order to dynamic determine what the relative close distance in local data space is, it is need to analyze the local data's distribution characteristics, such as the shape, size, density of local cluster. Since the cluster size of shared neighbors is pre-determined by user input parameter, k (the number of SNN), which select according to the type of target

data collection, then the local characteristics of the data gathering can be simplified to the representation of the local shape and density, where density can be measured by the average distance LAD (Local Average Distance) of all members of the shared neighbors. As the arbitrariness of local data distribution, the distribution shape measure can be reduced to two main aspects, i.e., local maximum distance LMD (Local Maximum Distance) and the local radial distance LRD (Local Radial Distance), the figure 2 give the illustrates of what LMD and LRD are. So we can get the definition of local data features as followings:

$$d_{LMD} = \max\{\text{distance}(p-p') \mid p \in C_{SNN}, p' \in C_{SNN}, p \neq p'\} \quad (1)$$

$$d_{LRD} = \max\{\text{distance}(p, \text{LineX}) \mid p \in C_{SNN}\} \quad (2)$$

$$d_{LAD} = \frac{2}{n(n-1)} \sum_{p \in C_{SNN}, p' \in C_{SNN}, p \neq p'} \text{distance}(p, p') \quad (3)$$

Where C_{SNN} is the shared nearest neighbors, n is the number of the data point in C_{SNN} , LineX is the line connected by two objects which have max distance in local cluster.

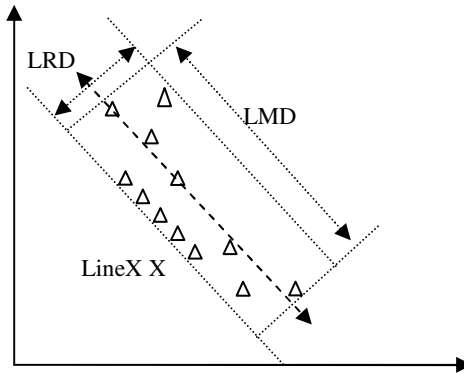


Fig. 2. Local characteristics measure

4.3 Improved LAC Based Clustering Algorithm

Among the previous analysis and definition, the local gathering features can be combined with JP algorithm to create a new clustering algorithm, i.e. improved LAC clustering algorithm. This algorithm is similar to JP and SNN density based clustering algorithms in mainly steps, however, it takes local data distribution into account, using LMD and LAD as a local dynamic threshold to control SNN similarity computation. The steps of improved LAC algorithm are shown in followings:

Identify the k nearest neighbors for each data object (the k data objects most similar to a given object, using a distance function to calculate the similarity).

Utilize LMD and LAD as dynamic control threshold to calculate the SNN similarity between pairs of data object and generate similarity graph.

Find the connected components (clusters) by applying a similarity threshold, and dynamic adjust the cluster membership at the same time.

Applying a similarity threshold to sparsify cluster graph can reduce compute complexity and improve algorithm efficiency in clusters finding. After graph sparsify, there are need a method to find and illustrate what kind cluster in graph. How to find connected components, here we give the pseudo code as followings:

```

While (exist data object not clustered) do
  Begin
    Random select a not clustered data point as seed and marked it as a new kind
    cluster member. Put seed in pool.
    While (not empty of seed pool) do
      Begin
        Fetch first seed from pool;
        Search similarity matrix finds all data objects have high similarity with
        fetched seed, marked them the same cluster, put them into seed pool at the
        same time.
      End
    End
  End
End

```

4.4 Experimental Results and Evaluation

The cluster evaluation is a very important part of cluster analysis, because almost every cluster algorithm will find clusters in a data set, even if that data set has no natural cluster structure. Since there are a number of different types of clusters-in some senses, each clustering algorithm defines its own types cluster to fit the destination data set. So design different clustering method to analysis different data objects, there are must have suitable clustering validation. The advantage of distance-based clustering is that distance is easy for computing and understanding. There are several choices for similarity definition between two sets of data points, as follows: [8]

$$Similarity_{rep}(C_i, C_j) = distance(rep_i, rep_j) \quad (4)$$

$$Similarity_{avg}(C_i, C_j) = \frac{1}{n_i \times n_j} \sum_{v_i \in C_i, v_j \in C_j} distance(v_i, v_j) \quad (5)$$

$$Similarity_{max}(C_i, C_j) = \max\{distance(v_i, v_j) | v_i \in C_i, v_j \in C_j\} \quad (6)$$

$$Similarity_{min}(C_i, C_j) = \min\{distance(v_i, v_j) | v_i \in C_i, v_j \in C_j\} \quad (7)$$

To test the effectiveness and accuracy of the proposed clustering method, the splice and synthetic dataset is adapted. In improved LAC algorithm evaluation experiment, total 150 data object being processed. Table 1 has shown the part data object.

Table 1. Part experiment data objects

No.	Standardize X	Standardize Y
1	0.7800872	0.180566049
2	0.61186137	0.43378935
3	0.65880533	0.19837862
4	0.33866129	0.23050593
5	0.89341246	0.92718452
6	0.22261332	0.69293582
7	0.52261332	0.02761889
...

With the same prerequisite, we run JP algorithm, SNN density method and improved LAC scheme on the experimental data collection, and the experimental result shown in table 2. Through compare with the difference algorithm's running result on statistical data, find the new presented method can get better aggregated clusters to deal with nature distributed data set and objects.

Table 2. The experiment result

Algorithms	Clusters Found	Average Similarity
SNN density	7	0.29
Jarvis-Patrick	6	0.32
LAC-SNN	4	0.35

4.5 Its Strength and Limitation

In some extend the JP method looks like brittle, due to its single link notion; the SNN density based algorithm maybe view small cluster as noise and discarded it, because it depend on ridged input parameters. In our view, both these problems rise from the overlook of clustering's local gathering features. To solve these problems there are need define an improved similarity measure based on the nearest neighbors of the objects. This similarity measure is not only limited to consider the direct nearest neighbors, but also can take into the neighbors' local configuration features. Furthermore, the suggested similarity measure fuzzifies the crisp decision criterion of the Jarvis-Patrick algorithm; and trade off against SNN density based method. Because improved LAC algorithm utilize a SNN similarity-based definition of a cluster with local dynamic threshold control, it is relatively resistant to noise and can handle clusters of arbitrary shapes and sizes, can deal with clusters of different density and natural distribution characteristics. Thus improved LAC not only can find many clusters that could not be found using K-means and DBSCAN, but also can get better analysis result than Jarvis-Patrick and SNN density algorithm, even if the clusters have widely varying or gradually changed densities. Even though widely fit able to different type data set, improved LAC scheme still exist some limitations, some time it can not get best representation of the local agglomerative characteristics in some type data space, so the representative method of local characteristics still need to be improved for some types destination data set.

5 Conclusions

In this paper we review the most widely used and successful SNN based clustering techniques and their related applications, summarize their shortcoming and limitation, give more detail and reasonable LAC's definition and introduced a new improved clustering algorithms based on LAC. The new algorithm overcome many of the challenges traditionally clustering algorithms, e.g., finding clusters in the presence of noise and outliers and finding clusters in data that has clusters of different shapes, sizes, and density, and finding more natural distribution clusters. The improved algorithm can give strong representative of local agglomerative characteristics of the target data set; can show flexible ability to fit sharply or gradually changed data density environments; can find relative small cluster but group tightly; can realize complete clustering and act outliers detection. De fact, the LAC clustering algorithm provides an effective tool for exploring groups of data.

References

1. Busygin, S., Prokopyev, O., Pardalos, P.M.: Biclustering in data mining. *Computers & Operations Research* 35, 2964–2987 (2008), <http://www.elsevier.com/locate/cor>
2. Almeida, J.A.S., Barbosa, L.M.S., Pais, A.A.C.C., Formosinho, S.J.: Improving hierarchical cluster analysis: A new method with outlier detection and automatic clustering. *Chemometrics and Intelligent Laboratory Systems* 87, 208–217 (2007)
3. Oyana, T.J.: A New-Fangled FES-k -Means Clustering Algorithm for Disease Discovery and Visual Analytics. *EURASIP Journal on Bioinformatics and Systems Biology* (2010)
4. Ertoz, L., Steinbach, M., Kumar, V.: A New Shared Nearest Neighbor Clustering Algorithm and its Applications. In: *Workshop on Clustering High Dimensional Data and its Applications at 2nd SIAM International Conference on Data Mining* (2002), <http://www.bibsonomy.org/bibtex/2ba0e3067111d2e3eea9a4d9fc995e36b/hotho> (2011)
5. Hu, T., Xiong, J., Zheng, G.: Similarity-based Combination of Multiple Clusterings. *International Journal of Computational Intelligence and Applications* 5(3), 351–369 (2005)
6. Tan, P.-N., Steinbach, M., Kumar, V.: *Introduction to Data Mining*. Pearson Education, 427–488 (2006)
7. Moreira, A., Santos, M.Y., Carneiro, S.: Density-based clustering algorithms–DBSCAN and SNN (2011), <http://ubicomp.algoritmi.uminho.pt/local/download/SNN&DBSCAN.pdf>
8. Qian, W.-N., Zhou, A.-Y.: Analyzing Popular Clustering Algorithms from Different Viewpoints. *Journal of Software* 13(8), 1382–1394 (2002)

Cluster Analysis Based on GAPSO Evolutionary Algorithm*

Tieqiao Huo¹, Junxi Zhang², and Xiaojun Wu^{1,3,**}

¹ College of Economics and Management, Xi'an University of Technology, Xi'an 710072, China

² Xi'an Aerotechnical College, Xi'an 710077, China

³ School of Computer Science, Shaanxi Normal University, Xi'an 710062, China
jinqisu@163.com

Abstract. Cluster Analysis which plays an important role in Data Mining, is widely used. It has important value both in theory and application. Considering the stability of the Genetic Algorithm and the local searching capability of Particle Swarm Optimization in clustering, those two algorithms are combined. Particle Swarm Optimization operators are implemented after the crossover and mutation operators, and GA-PSO clustering algorithm is put forwarded. Simulation results are given to illustrate the stability and convergence of the proposed method. GA-PSO is proved to be easier to carry out, faster to converge and more stable than other methods.

Keywords: Data Mining, cluster, Genetic Algorithm, Particle Swarm Optimization, GAPSO.

1 Introduction

Data Mining is to nugget out potential, hidden useful knowledge and information from abundant, incomplete, noisy, fuzzy and random practical data. With the development of information technology and artificial intelligent, lots of efficient data mining methods are put forwarded and widely used. It is becoming a hotspot in computer science [1].

Cluster Analysis is the core technology of Data Mining. It attempts to partition a dataset into a meaningful set of mutually exclusive clusters according to similarity of data in order to make the data more similar within group and more diverse between groups[2]. The objects are partitioned by certain requirement and rules. During that process, there is no prior knowledge, no guiding. Similarity between objects is the only partition rule. That belongs to Unsupervised Classification [3][4].

Clustering technique can be divided into 7 categories: hierarchical cluster, partitioning cluster, density-based cluster, grid-based cluster, character attribute joint cluster, multi-dimensional data cluster and NN cluster[2]. K-means algorithm in

* Supported by the Fundamental Research Funds for the Central Universities(GK201002005), Xi'an Science and Technology Innovation Programming(CXY1016-2), Shaan Xi Industrial research program(No. 2009K09-21).

** Corresponding author.

partitioning cluster is the most widely used, but the traditional k-means cluster algorithm has its inherent limitations: ①Random initialization could lead to different clustering results, even no result. ②The algorithm is based on objective function, and usually take the Gradient method to solve problem. As the Gradient method searched along the direction of energy decreasing, that makes the algorithm get into local optimum, and sensitive to isolated points.

Genetic Algorithm performs well in global searching. Reference [5] introduced Genetic Algorithm in cluster analysis as follows: a chromosomal is divided into N parts, and each part is corresponding to a type of dataset. The adaptive function is defined as the Euclidean distance between data. The initial population is randomly obtained. Generating operation is similar to traditional genetic algorithm. This method could find out the global optimal solution, not being effected by the isolated points. But when confronted with abundant data and large amount of types, the computation time is too long and the convergence speed is not satisfying. But there is still lots of researching fruits using GA in clustering[6][7].

PSO has good convergence capability. Reference [8] combined it with GA, and solved genetic programming well. Pure PSO is also useful in special cluster[9][10], but due to its disadvantage in global searching capability[11], and uncertain operators related to practical problems, there are some limitations using PSO in cluster analysis.

On the basis of reference [5], GAPSO cluster algorithm is put forwarded combining with PSO. It takes both the stability of the Genetic Algorithm and the local searching capability of Particle Swarm Optimization. The result proved that this method outperformed the GA and pure PSO in clustering efficiency.

2 PSO Algorithm

PSO was proposed by Kennedy and Eberhart in 1995[12]. Birds searching for food are abstracted as the solution of a problem. So searching for optimum in the searching space is transformed to a problem of birds searching for food. The solution in searching space is called “particle”. All the particles have an optimized function to measure the distance of its current position to the global position of swarm, called “fitness”. Except for “position”, each particle has velocity information to denote flying direction and distance. The position of particles is adjusted according to the best particle and its own best position.

PSO is an evolutionary technique through individual improvement plus population cooperation and competition which is based on simulation of simplified social models. PSO starts with a random initialization of a population of individuals in the searching space. Then the optimum is obtained through iteration. In each iteration step, particles are updated through running after two “extremums”. One is the best local solution, and the other is the best global solution. Otherwise part of the population can be used as the neighbor of particles, and the extremum in the neighbors is the best local solution.

As integrated information of local and global optimum is added in PSO, the update of position is more targeted. So PSO has high speed of convergence.

In the improvement by Eberhart R.A in 1998, the position of “particle” is updated according to the following formula[13]:

$$\begin{cases} V_i = w * V_i + c_1 * r_1 * (X_{ip} - X_i) + c_2 * r_2 * (X_g - X_i) \\ X_i = X_i + a * V_i \end{cases} \quad (1)$$

where V_i is named flying speed of particle at time i , X_i is the position, X_g is called the best global solution, X_{ip} is called the best local solution, w is called the inertia factor, r_1 and r_2 are two independent random numbers. C_1 and C_2 control the optimization assignment between local and global optimum.

3 GAPSO Cluster Algorithm

There are 3 points different from traditional GA in GAPSO:

- (1) When calculating the fitness of individuals, the positions of the best global particles should be supervised and recorded.
- (2) PSO is processed after executing GA operators. Generally speaking, one in two method or one in three method is selected.
- (3) GAPSO cluster algorithm needs determining the population size, selecting intensity, probability of crossover and mutation, and also the optimizing proportion of PSO and related PSO parameters.

GA performs well in global searching, but from the viewpoint of crossover and mutation, solutions are adjusted significantly. Especially when the obtained solution is near by the best solution, its searching capability would descend in stead. Concerning the high speed of convergence of PSO, PSO is added after crossover and mutation of GA, the flow chart is as follows:

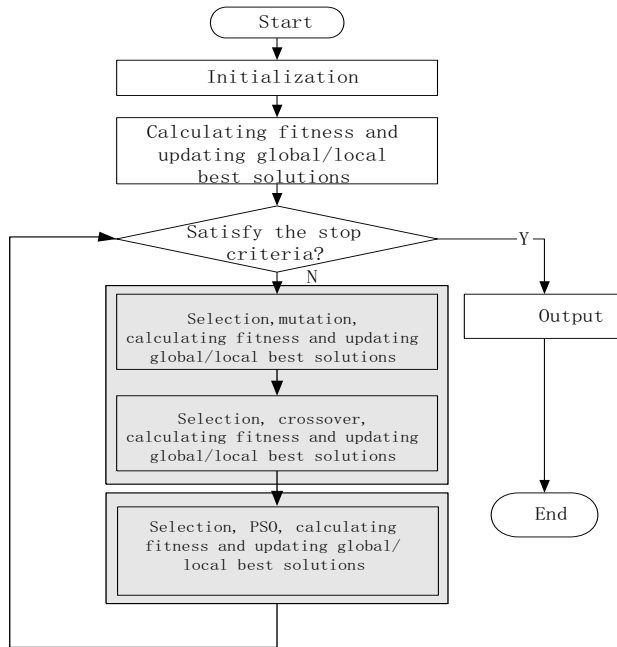


Fig. 1. Flow chart of GAPSO Cluster

3.1 Coding and Initial Generation

According to the method in reference [5], coding is implemented by the cluster centroid. Suppose there are n samples $X_i(i=1,2,\dots,n)$, m cluster centroids $P_j(j=1,2,\dots,m)$, then that m cluster centroids are denoted by a chromosome: $S=P_1P_2\dots P_m$. partition of cluster is determined by Nearest-Neighbor Algorithm. If the distance from X_i to P_j satisfies the following equation, then X_i is partitioned to cluster centroid j .

$$D(X_i, p_j) = \min_{k=1,2,\dots,m} (D(X_i, P_k)) \quad (2)$$

where $D(x_i, p_j)$ indicates the distance from X_i to P_j , which could be Minkowski distance, Euclid distance, Manhattan distance, in this experiment, we choose Euclid distance.

3.2 Calculation of the Fitness

The distance between X_i and the cluster centroid is regarded as fitness of the algorithm.

Suppose there are n given objects to be partitioned, based on formula (2), C_i denotes the cluster centroid of object I , $D(X_i, C_i)$ denotes the distance between X_i and C_i . Fitness can be calculated in the following way:

$$fit = \sum_{i=1}^n D(X_i, C_i) \quad (3)$$

which is a universal description. In this experiment, the objects are in two-dimensional space, and sum of squares of the distance between objects and cluster centroid is considered as the fitness.

Owing to the adding of PSO, as the fitness is updated after crossover, mutation and PSO were implemented, the best local and global solutions should also be updated.

3.3 Genetic Operator

Mutation includes one in two selection method and single point stochastic mutation; Crossover includes two in three selection method and two-point crossover.

3.4 PSO Operator

The core technique of PSO is the update of solution. According to formula (1), it is concluded that PSO refers to both the best local and global solution, so the searching is more sensitive. But there is still a problem to use formula (1) directly: the design of PSO considered the method of boosting its global search capability, so inertia factor and constraint factor are added. To GAPSO, the determination of parameters brings difficulty to implementation of the algorithm, and decreases stability and universal utility. The experiment also proved that parameters of PSO are very sensitive to population size and type of solution. So formula (1) is modified as follows:

$$X_i = X_i + a * (r * (X_{ip} - X_i) + (1 - r)(X_g - X_i)) \tag{4}$$

where a is the distribution coefficient, commonly choosing $a=2$. r is a random number, $r \in (0, 1)$, X_i represents one of the solutions. X_{ip} denotes the optimal solution in the calculating process. X_g is the global optimal solution.

In formula (4), both the global optimal solution and the local optimal solution are considered. That promotes the local searching capability of the algorithm, and reduces undetermined coefficients, so GAPSO cluster algorithm performs better in stability and universal utility.

4 Experiment Conclusion

10 datasets in two-dimensional space are selected in this experiment. Each group is implemented by GA, PSO and GAPSO in turns, the average fitness are calculated on the same iteration step. Fig.2 and Fig.3 summarize the space distribution and partition of dataset 1 and dataset 2, where “+”denotes the optimal cluster centroid:

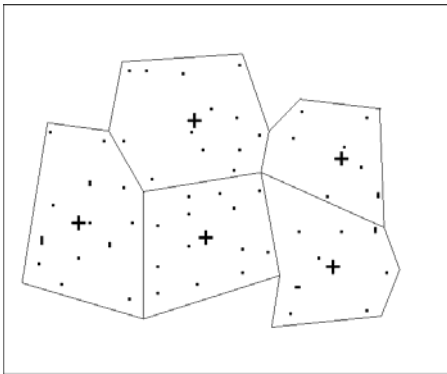


Fig. 2. Clustering result of dataset 1

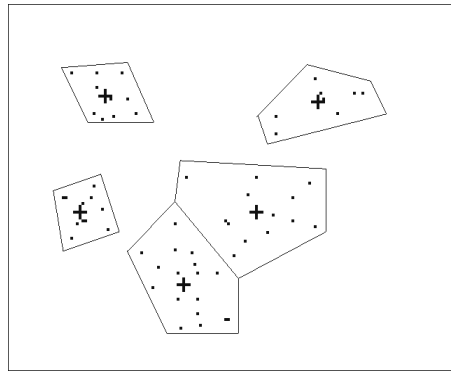


Fig. 3. Clustering result of dataset 2

Compared with GA, PSO and GAPSO have obviously modified the convergence on average fitness in those 10 datasets. Within the observational range, GA can't converge to the optimal solution in certain dataset, but PSO and GAPSO usually converge to the same optimal solution. In comparison between the 10 datasets implemented by PSO and GAPSO respectively, at the first stage, they converge rapidly. But after several iteration cycles, GAPSO could converges to an optimal solution steadily, however PSO need longer time on iteration or even couldn't converge to the optimal solution that GAPSO has achieved. So GAPSO cluster algorithm has higher convergence efficiency than GA, and better stability than PSO.

Fig.4. shows the comparison between GA and GAPSO by dataset 1, it is detected that there are large differences between GA and GAPSO, even after iteration step 300, GA could only achieve the level of GAPSO. Fig.5. shows the comparison between PSO and GAPSO by dataset 1. The result proved that GAPSO has better stability than

pure PSO. Fig.6. and Fig.7. are comparative convergence curve by dataset 2. It also proved that GAPSO has higher convergence efficiency than GA and better stability than PSO.

In the algorithm evaluation, iteration time is used to compare convergence. From the calculating process, GA, PSO and GAPSO are different in calculating efficiency. But in practical, there are little differences. So the difference wasn't considered in the experiment.

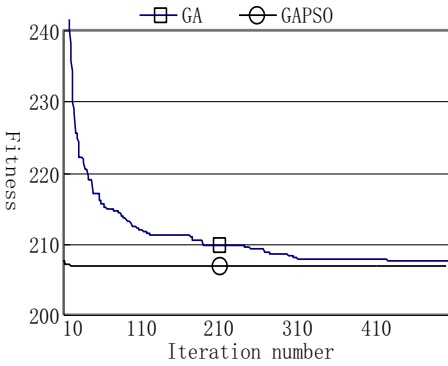


Fig. 4. GA and GAPSO convergence curve of dataset 1

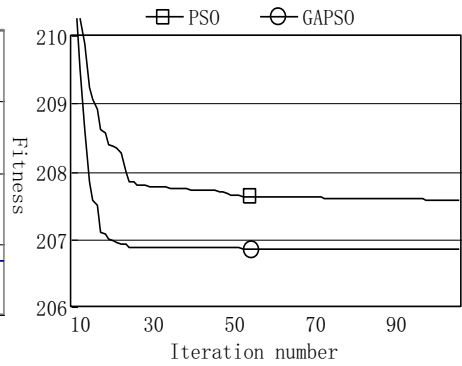


Fig. 5. PSO and GAPSO convergence curve of dataset 1

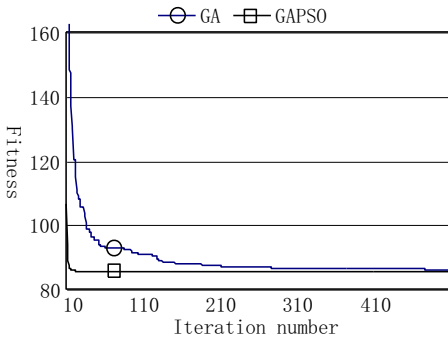


Fig. 6. GA and GAPSO convergence curve of dataset 2

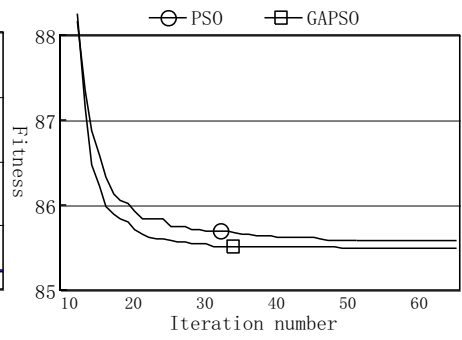


Fig. 7. PSO and GAPSO convergence curve of dataset 2

5 Conclusion

Cluster analysis is the core technique of data mining, and it is valuable to modern times as information technique is widely used. Traditional cluster algorithm has the characteristic of unstable and is very sensitive to isolated points. GA-based cluster analysis is another researching focus, but it sacrifices convergence for stability. PSO emphasizes convergence but its coefficients are difficult to determine and is not stable. This paper added PSO to GA, considering both of their characteristics, and put forward GAPSO cluster algorithm. This method takes both the stability and universal utility of GA and high convergence efficiency of PSO.

In this experiment, several runs were performed by the 10 datasets. Through the analysis of average fitness, GAPSO cluster algorithm outperforms GA on convergence and PSO on stability.

References

1. Yang, J.-j., Deng, H.-w., Zi, T.: Cluster Analysis Based on the Chaos Immune Evolutionary Algorithm. *Journal of Computer Science* 35(8), 154–157 (2008)
2. Yuan, J., Zhang, Z.-y., Qiu, P.-l., Zhou, D.-f.: Clustering Algorithms Used in Data Mining. *Journal of Electronics and Information Technology* (4), 655–660 (2005)
3. Han, J.W., Kamber, M.: *Data mining: concepts and techniques*. Morgan Kaufmann, San Francisco (2000)
4. Tan, P.-N., Steinbach, M., Kumar, V.: *Data Mining Instruction*. Posts & Telecom Press, Beijing (May 2006)
5. Ujjwal, M., Sanghamitra, B.: Genetic Algorithm Based Clustering Technique. *Pattern Recognition* 33(9), 1455 (2000)
6. Babu, G.P., Murty, M.N.: A near-optimal initial seed value selection in K-means algorithm using a genetic algorithm. *Pattern Recogn. Lett.* 12(10), 763–769 (1993)
7. Bi, L., Yong, Z.-z.: Clustering with a Modified Genetically Optimized Approach. *Journal of Circuits and Systems* (3), 96–99 (2002)
8. Wu, X.-j., Xue, H.-f., Min, L., Lan, Z.-l.: A new programming of mixed genetic algorithm with particle swarm optimization. *Journal of Northwest University(Natural Science Edition)* (1) (2005)
9. Qiang, L., Jinshou, Y.: Fuzzy C-Means Algorithm Based on Particle Swarm Optimizer for Optimal Clustering and the Application for the Optimization of Acrylonitrile Reactor. *Computer Engineering and Applications* 22, 211–214 (2005)
10. Ni, L., Duan, C., Jia, C.-l.: Hybrid particle swarm optimization algorithm based on differential evolution for project scheduling problems. *Application Research of Computers* 28(4), 1286–1289 (2011)
11. Aiguo, L., Zheng, Q., Fumin, B., Shengping, H.: Particle Swarm Optimization Algorithms. *Computer Engineering and Applications* (21), 1–3 (2002)
12. Kennedy, J., Eberhart, R.C.: Particle Swarm Optimization. In: *Proc. IEEE int'l Conf. On Neural Network*, vol. (4), pp. 1942–1948. IEEE service center, Piscataway (1995)
13. Shi, Y., Eberhart, R.: A modified particle swarm optimizer. *IEEE World Congress on Computational Intelligence*, 69–73 (1998)

A Model of a GEP-Based Text Clustering on Counter Propagation Networks

Jin'guang Luo, Chang'an Yuan, and Jinkun Luo

School of Computer and Information Engineering, Guangxi Teachers Education University,
Nanning, 530023, China
yca@gxte.edu.cn

Abstract. In this paper, we present a model of a GEP-based text clustering research on counter propagation networks. The idea of the model is that it optimize the link weight vector by using the advantage of the GEP(Gene Expression Programming, GEP) in combinatorial optimization. We investigate how the value of weight in the network affect the performance of the text clustering by comparing it to a based on genetic algorithm and SOM network and the method of the traditional CPN(Counter Propagation Networks, CPN). Furthermore, we improve and optimize the weight in the CPN network by the method of GEP, thus raise the quality of the text clustering in the network. Finally, in this paper, we demonstrated the validity and superiority of the presented model.

Keywords: Text Clustering, Weight Optimization, Neural Network, Gene Expression Programming, Counter Propagation Networks.

1 Introduction and Related Works

Clustering algorithms generally used are: AHP, segmentation method, grid-based methods, methods based on density etc. which there are two more classical methods—K-means and SOM-based neural networks, and the improved algorithms for them[1,2,3,4,5]. Literature[1] present a k-means text clustering algorithm based on similar centroid, literature[4] present a text clustering method based on genetic algorithm and SOM network. While these two improved methods would obtain some text clustering efficiency, but also own flaw. Such as, the lower convergence and maybe appear "dead neurons".

In the traditional CPN[6] (Counter Propagation Networks, CPN) network algorithm, the difference small pseudo-random numbers are used generally as the link weight initial value. As in most cases, the distribution of the input vectors in the sample set is uneven, and these small pseudo-random numbers distribute uniformly in a high dimensional "sphere", in which case, uneven input vectors would cause the weight vectors trend and gather around a certain part of the "sphere". While the most other weight vectors are so too far from the input vectors that could not get the matching neuron. So, there will appear "dead neurons", which lead to waste that group link weights of this neuron and corresponding link weights and what in the Grossberg.

The literature [7] present a text classification algorithm based on GEP(Gene Expression Programming, GEP), which is an algorithm based on composite discriminate model with GEP. From the literature, the author also presented that its convergence speed was slower and have a long time, so, base on which, from overcoming the “dead neurons” and slowing convergence speed in the literature [4, 5, 7], we present a model of a GEP-CPN-based text clustering algorithm (GCTCA). On the one hand, the CPN consists of two algorithms, one is SOM algorithm (or Kohonen algorithm), other is Outer algorithm, so which include self-organizing capacity, its training speed is more faster than BP network; on the other hand, through introducing GEP[8], which is better than the genetic algorithm[9], so it have great superiority on the impact of solving the initial state of the link weight value in network from convergence in network. The experimental results also show that the new algorithm could avoid appearing “dead neurons” that could waste the weight of the corresponding neuron, and get better result.

As CPN network has two layers, one layer is a Kohonen that have classification functions, and the other is Grossberg layer that have transformation function, so the network can be applied to the classification problem of expression, decision support, planning, etc., while the literature [10, 11, 12] search in the following fields: predict the respective categories of information, web text extraction and structural identification etc.

The reminder of the paper is organized as follows: in the second section, we present our model of a GEP-based text clustering on CPN(Counter Propagation Networks, CPN).Third, this is a analysis of the set and the result in the whole experiment. Finally, we conclude with a summary of the work and look into the future work.

2 Model Descriptions and Analysis

In the traditional algorithm, initialize the link weight vector in the Kohonen, the random (0,1) is the initial value, which appear the phenomenon of section 1, appear the “dead neurons” and then lead to waste cost. And if the weight vector responding to a neuron get the match that is more aggregate, thus the clustering efficiency the training result is not obvious, even to waste time. So, in this paper, introduce the GEP algorithm, which is similar to GA and GP on the main steps, but it only have larger contributions in function discovery[9], association rules[13] etc. but also it have more versatility in combinatorial optimization[8], so the new algorithm optimize the initialization of link weight vector in Kohonen with the GEP’s advantage.

2.1 The Model Description

In order to avoid the inconsistent between the initial value of link weight and input vector directly—some weight vectors get more input matching, even to not, and others get less. For that, the new model optimize this phenomenon with the capacity of the GEP’s global optimization search for raising to the accuracy of the CPN network and overcome the shortcoming of the traditional method.

The main steps of the new algorithm in this paper are: first, the structure of text is set by using the VSM[14], the text information is reduced by using conventional methods, and then remove duplicate or non-words and get the text feature set; the second step is: a set of text features formed the input vector of the CPN network, using GEP algorithm optimize the initial value of link weight between input vector and Kohonen, its objective is: according to the actual distribution of the input vector to set the corresponding weight vectors of the each neurons in Kohonen, algorithm 1 implements this function; the third step is: establish the new CPN model (GCTCA) and then train it for learning, at the same time, it collects clustering results. This clustering model is essentially a global optimization algorithm using GEP's adaptive search function to solve problem that is inconsistent between distribution of the initialed vector in Kohonen layer and input vector, to jump out the phenomenon of local aggregation as much as possible. The specific implementation steps in the next section.

Algorithm 1. Optimize link weight value with GEP algorithm

input : text vectors;

output : $Best (K_j)$

1) Read() ;

2) Compute the weight value responding to the m “winner” with equation 1;and others are computed;

3) $W^{(new)} = W^{(old)} + \alpha(X - W^{(old)})$;

4) If W_{mj} is only the same value, end; // W_{mj} is the best weight;

 Else next step;

5) genetic manipulation produce the next generation:

 Selection; Replication; Mutation;

 Transposition;//this mainly transpose the domain of W_j , literature[7]have described in detail the methods for the domain transposition;

 Recombination; Inversion;

6) Return 2);

2.2 Achievement and Analysis of Model

The design of CPN network is mainly the choice of the network topology, in this paper, the number of the input layer is 10, and the output is 2. According to relevant researchers [15] presented some empirical guidance, set the competitive layer's number is 8. From the section 3.2, which shows that, the main idea of GCTCA algorithm, at the early stage, few neurons are as the “winner”, then adjust the corresponding weight vector of these “winners”, the equation 4 get the appropriate adjusted function (it is said that the fitness function in GEP algorithm), which shows that each neurons adjusted would be compared, and then get a neuron that have the best value as the “winner”.

$$Best(W_{mj}) = Best(W_{mj} + \alpha_m(X - W_{mj}))$$

Where α_m the random value in $(0, 1)$, m is the number of "winner" in the early stage.

After establishing the fitness function, algorithm must code firstly, which code the input vectors and the link weight using the GEP's encoding. The input vectors is $(X_1, X_2, \dots, X_{10})$, link weight value vector is:

$$W_j = (w_{j1}, w_{j2}, \dots, w_{ji}), 1 \leq j \leq 8, 1 \leq i \leq 10$$

Because of the advantage of GEP algorithm in the convergence, the convergence speed is faster than traditional CPN network. Algorithm 2 instruct the process of the whole model. Based on the algorithm 2 can be seen that, in the beginning of algorithm, it determine the m neurons in the excited state firstly, and then using the evolution of GEP to get an input vector only responding one neuron excited, and the traditional CPN algorithm, at the beginning, it does not optimize the initial value so that it maybe fall into clustering the local range, so it doesn't reach the goal optimum for text clustering. Instead, in the new algorithm, it solves the initial distribution of the weight vector consistent with the input vector distribution at the beginning. So, in theory, the algorithm can obtain better clustering results, it can also be reduced because of "too much from the far "and result in waste of some neurons, increase network utilization.

Algorithm 2. GCTCA algorithm

input : $\langle X_1, X_2, X_3, \dots, X_{10}, j \rangle$

output: reflect the situation of the text clustering

- 1) Initialize CPN Network;
- 2) Set $m=3$, $\text{random}(W_j) \in (0,1)$;
- 3) Initialize $\langle X_1, X_2, X_3, \dots, X_{10}, W_j \rangle$;
- 4) Call algorithm 1;
- 5) Processing $Best(K_j)$ to units;

For 1 to j do {

- 6) Update V using equation 5:

$$V_j^{(new)} = V_j^{(old)} + \alpha(Y - V_j^{(old)}) \quad (5)$$

}

- 7) Output the result;
-

3 The Result and Analysis of the Experiment

The experimental dataset is the text data that are taken from the UCI KDD Archive (<http://kdd.ics.uci.edu/databases/20newsgroups/20newsgroups.html>) mini newsgroups and 20 News Groups, where mini newsgroups is a subset composed of 100 articles

from each newsgroup,20 Newsgroups collect 20,000 documents of 20 Usenet newsgroups. In the text clustering, commonly, the evaluation is the retrieval indicators that are the most classic in the information field: recall, precision-measure, the specific formula of which are shown in the literature[16].

3.1 Availability

In the following, this paper make an experiment about the availability of the CPN on the clustering. The purpose is to prove that CPN can be used for cluser analysis. The dataset is the mini newsgroups from the UCI KDD Archive. First, the training set is 30 articles that are extracted randomly from each newsgroup; the remaining is the test sample. The result of traditional CPN algorithm is shown on table 1.

From the experimental results, it is not very satisfactory results by training the test set in the network, which has low percentage of correct clustering. And from the experimental observation, if a data and other data have some common ground, the network did not have satisfactory results. Such as: “talk.religion.misc” and “talk.politics.misc”, one is “religion”, the other is “politics”, but it is described as the “misc”, when the output is ariseded in the network, these two categories may run into "wrong" target, so it is easy to fall into a "trap", thus the result of their clustering has less effective.

Table 1. The result of clustering

Category	correct ratio	Category	correct ratio
alt.atheism	79.25%	rec.sport.baseball	77.04%
comp.graphics	80.12%	rec.sport.hockey	80.39%
comp.os.ms-windows.misc	78.10%	sci.crypt	80.23%
comp.sys.ibm.pc.hardware	79.41%	sci.electronics	76.45%
comp.sys.mac.hardware	80.14%	sci.med	79.05%
comp.windows.x	81.05%	sci.space	79.77%
misc.forsale	80.41%	soc.religion.christian	80.01%
rec.autos	79.04%	talk.politics.guns	79.78%
rec.motorcycles	79.99%	talk.politics.mideast	73.98%
talk.religion.misc	75.23%	talk.politics.misc	70.23%

3.2 Results and Analysis

The experimental operating environment is: CPU: Intel(R) Core(TM)2 Duo CPU 5670 @1.80 GHz with 1GB memory, run in Microsoft Windows XP Professional. Population operating parameters are: number of run is 10000, evolution generation is 500, population size is 100, the head length of gene is 40, mutation rate is 0.05, single gene recombination rate is 0.1, double gene recombination rate is 0.1, mutation rate of weight is 0.002, IS transposition rate is 0.1.In the training network lab, we suppose

there are 2 or 3 neurons in the excited state, which will get to a better convergence and clustering result. And these two experiments are on the Visual Studio 2008 with C#.

This paper finishes two experiments, the first uses the mini newsgroups to prove the convergence of the model algorithm. The second uses 600 articles randomly from each newsgroup, which are the set of sample. Before starting the experiment, we use VSM [14] to set the structure of the text set, and then apply these text vectors to the text clustering with the GCTCA algorithm.

3.2.1 Experiment 1

In order to evaluate the new algorithm in convergence, the first lab compare GCTCA algorithm to k-means by literature[1] and the GSTCA by the literature[4]. The result shows in figure 1.

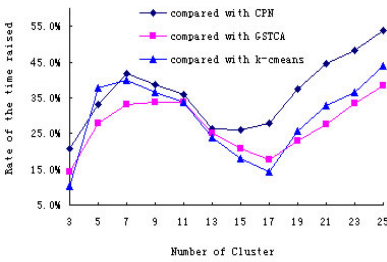


Fig. 1. GCTCA compared with CPN, GSTCA and K-means

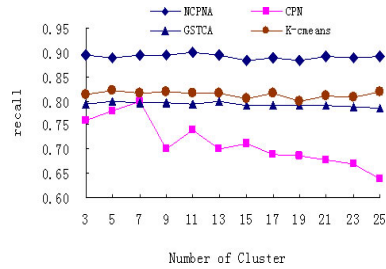


Fig. 2. Comparison of Recall

Figure 1 illustrates the convergence of GCTCA is better than CPN, GSTCA and K-means algorithm, especially CPN. Where, X-axis is the number of cluster, Y-axis is the rate of the time-raised. With the number increase, the convergence time of each algorithm occurrence obvious differences. And in the k-means algorithm, which is only overcome the shortcoming of the initial centre sensitive, and did not improve the sequence of input sample impact the result. The two other algorithms—GSTCA and CPN—is because of the weight vector can impact the consistency with the imputed vector, which can lead to appears “dead neuron”, so this will waste cost of the time. The new model can better than others from run time, because it optimizes the weight vector in the Kohonen network for improving the consistency between the initial weight vector and the input vector. The GCTCA algorithm increase to average of about 31.01% than other algorithms in the experiment.

3.2.2 Experiment 2

Figure 2, 3 and 4 show that the performance with recall, precision, F-measure, by compare the CPN and the GSTCA from the literature [4] and the algorithm—K-means—from the literature [1].

Figure 2 and 3 show that the change of the recall and precision in the four methods with the number of cluster. From the figure we know that the GCTCA and GSTCA always maintain a certain value, volatility is more stable. And the CPN is less stable,

especially, on the recall rate, with number increase, it become smaller and less stable, because when initial the link weight in the Kohonen, generated weight vector randomly did not consist with the actual input vector. Relative to the GCTCA and GSTCA, on a certain extent, they avoid to failing into cluster in local range for achieving optimal, so this get the better clustering accuracy and stability, the volatility of which is little big. Figure 4 describes the situation of the comprehensive quality of the three algorithms, from the figure we can find no matter how the cluster times, the clustering quality of the new algorithm is better than other two.

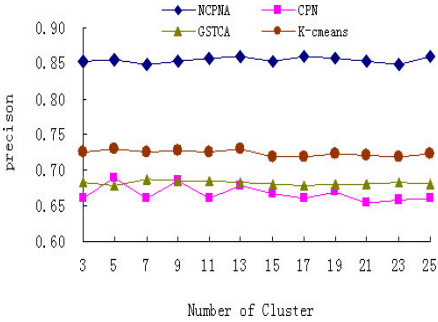


Fig. 3. Comparison of Precision

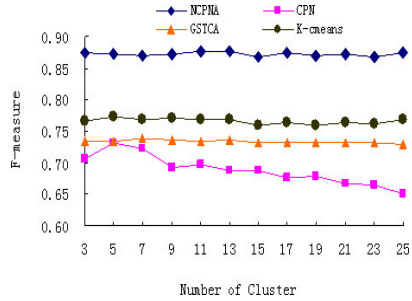


Fig. 4. Comparison of F-measure

From the figures we know that GCTCA algorithm is better than GSTCA and CPN in the clustering accuracy. The stable of GSTCA is better, but it optimize the initial weight with GA (genetic algorithm), and the GA have the shortcoming in capacity on finding complex data, so discount in the clustering accuracy. From the comparison of recall, the GCTCA raises to average of about 19.97%(CPN), 11.06%(GSTCA), 8.76%(K-cmeans); About the precison, raises to about 21.96%(CPN), 20.20% (GSTCA), 15.30%(k-cmeans); In the comparison of F-measure, raises to about 21.07%(CPN), 15.97%(GSTCA), 12.22%(k-cmeans).

4 Conclusion and Future Works

This paper research the CPN neural network, and as a basis of which, optimize the initial link weight vector in Kohonen for achieving better text clustering. And advance the convergence, which can be able to avoid the network fail into the local best. Furthermore we analysis the new algorithm by using two experiments, the experimental result show that GCTCA algorithm get the better clustering efficiency in clustering discovery. The m of a number of neurons excited state in this paper, which is obtained by several experiments, if $m > 3$, this network split the class seriously, some text may be split into other class but the same, or split into two class. The next work is the reduction of the complex data input on the basis of this paper and not impact the convergence speed, to achieve the text clustering in the complex data for increasing the convergence speed and apply it to the trust terminal user or intrusion detection.

Acknowledgments. This research is sponsored by the Institute of Science Computing and Intelligent Information Processing, and supported by the following projects: (1) the National Natural Science Foundation of China (Grant No.60763012); (2) the Natural Science Key Foundation of Guangxi (Grant No. 2011GXNSFD018025)); (3) the Innovation Project of GuangXi Graduate Education (Grant No. 2010106030774M02).

References

1. Xu, H.-j., Liu, Y.-y., et al.: K-means text clustering algorithm based on similar centroid. *Computer Engineering and Design* 31(8), 1802–1805 (2010)
2. Kohonen, T.: *Self-Organizing Maps*, 3rd edn. Springer, Heidelberg (2001)
3. Han, J., Kamber, M.: *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, San Francisco (2000)
4. Xiao, Q., Yuan, C.-a.: Text clustering method based on genetic algorithm and SOM network. *Computer Applications* 28(3), 757–760 (2008)
5. Guo, T., Chen, H.-g.: Text Clustering Method Based on BBS Hot Topics Discovery. *Computer Engineering* 36(7), 79–81 (2010)
6. Wasserman, P. D.: *Neural computing: theory and practice*. Van Nostrand Reinhold (1989)
7. Qin, X., Yuan, C., Peng, Y., Ding, C.: A GEP-Based Text Classification Algorithm. *Journal of Computational Information Systems, Journal of Information & Computational Science* 6(3), 1303–1309 (2009)
8. Ferreira, C.: Gene expression programming: a new adaptive algorithm for solving problems. *Complex System* 13(2), 87–129 (2001)
9. Yuan, C.-a., Tan, C.-j., Zuo, J., etc: Function Mining Based on Gene Expression Programming—Convergence Analysis and Remnant guided Evolution Algorithm. *Journal of Sichuan University (Engineering Science Edition)* 36(6), 100–105 (2004)
10. Na, T., Zhang, J.-g.: Prediction of the customer value based on rough set and CPN network. *Statistics and Decision* (5), 155–156 (2008)
11. Chen, J.-w., Peng, Z.: Study on Web text extraction based on CPN network. *New Technology of Library and Information Service* 11, 65–71 (2008)
12. Gao, H.-s., Ye, T.-q.: Counter propagation Network-Based analysis for structural identification. *Microcomputer Development* (2), 28–29 (1999)
13. Zuo, J., Tang, C., Zhang, T.: Mining predicate association rule by gene expression programming. In: Meng, X., Su, J., Wang, Y. (eds.) *WAIM 2002*. LNCS, vol. 2419, pp. 92–103. Springer, Heidelberg (2002)
14. Salton, G., Yang, A.C.: A vector space model for automatic indexing. *Communications of the ACM* 18(11), 613–620 (1975)
15. Heaton, J.: *Java Neural Networks* (December 24, 2007). , <http://www.heatonresearch.com/article//5/page1.html>
16. Ayad, H., Kamel, M.: Topic discovery from text using aggregation of different clustering methods. In: Cohen, R., Spencer, B. (eds.) *Canadian AI 2002*. LNCS (LNAI), vol. 2338, pp. 161–175. Springer, Heidelberg (2002)

The Application of Data Mining Technology in Analysis of the Relation between Majors, Industries and Positions

Xiaoguo Wang and Lin Sun

Electronic Information Engineering Institute, Tongji University,
4800 Caoan Road, Jiading District, Shanghai, China, 201804
xiaoguoawang@tongji.edu.cn, sunlin_777@163.com

Abstract. In the context of Prediction System of University Major Setting Research Project, for the machinery manufacturing industry, we study for the association rules model of the relation between majors and positions. We design a set of methods to discover this model, achieve this model with existing data and analyze the practical significance of this model. This association rules model provides a useful exploration to university major settings and employment trend analysis.

Keywords: Association rules, visualization, the relation between majors, industries and positions.

1 Introduction

The correlation between majors, industries and positions has a great significance to the major setting and employment trends. The research on the model of the relation between majors, industries and positions is also an essential part of the Prediction System of University Major Setting Research Project.

The main job of this paper is discovering the association rules model between positions and majors for a specific industry, using association rule mining and visualization technology. The model can provide a scientific guidance to the university major setting and employment trend analysis.

2 General Design

2.1 Data Collection

The scale of China's machinery manufacturing industry has a leading position in the world. In this paper, we focus on this industry and study the correlation between majors and positions. According to the research needs, combined with the actual situation of the research enterprise, the project members and business executives

design the questionnaire. Then we investigate for the staff in a large machinery manufacturing enterprise and get their education background and current position information as our mining data.

2.2 Data Preprocessing

Data Cleaning. The main task of data cleaning is to eliminate redundant data and noise data, eliminate duplicate or invalid records. Some data records don't have the complete key attribute. That type of records is invalid and we should give these records a further research. On the basis of the research result, we can decide to amend the records or delete them; there are also some records with fuzzy information. That type of records should be given recognition and re-fill.

Data Standardization. The main task of data standardization is to establish the correspondence between Position Classification Dictionary of People's Republic of China and the performance standard of position classification in the enterprise. Based on that correspondence rules, we can design a performance standard of position classification for the Prediction System of University Major Setting Research Project and use it to carry out the standardization of position information. The process is shown in Fig. 1.

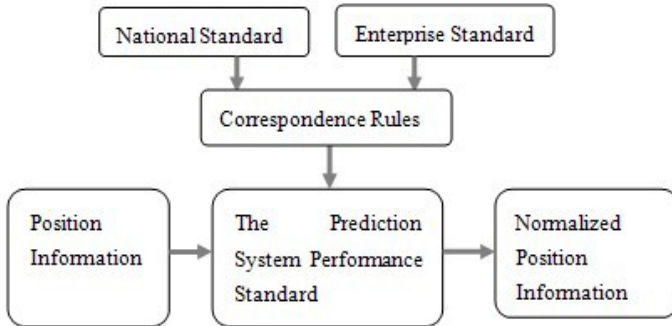


Fig. 1. The process of position information standardization

Data Transformation. Data transformation is an important part of data preprocessing. When the data source can't directly meet the data form that data mining algorithms require, we need to convert them. For example, gender, type of graduate school, educational level and some other attributes need to be converted to Integer type, according to the needs of association rules mining.

2.3 Algorithm Selection

Apriori is a classical algorithm, usually for mining single-dimensional and Boolean association rules in the transaction database. Based on the characteristics of our

mining task, we choose a multi-dimensional improved Apriori algorithm[1], this method is applicable to multi-dimensional association rule mining.

The multi-dimensional improved Apriori algorithm[1] is also separated into two steps. Firstly, search for the frequent item sets. Secondly, deduce the association rules from the frequent item sets. The different part is that we need to quantify the value of the properties before searching for the frequent item sets, using interval instead of the numerical value.

2.4 Visualization

Visualization provides a strong support to association rule mining process. To allow users to participate the mining process, we realize the visualization of mining process and mining results in this paper. Users can choose the properties involved in mining and set the minimum support and minimum confidence, according to their needs.

In summary, we mainly focus on four aspects of the association rules model: data collection, data preprocessing, the association rules model determination, interpretation and analysis for the model. The method flow chart is shown in Fig. 2.

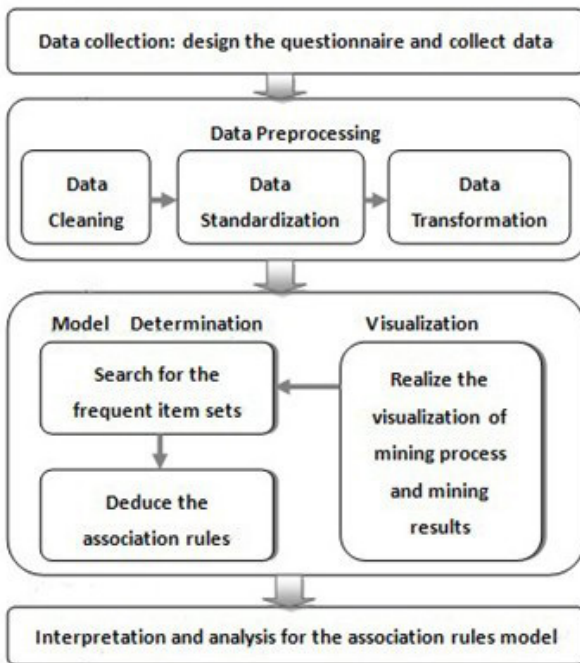


Fig. 2. The method flow chart for discovering the association rules model

3 The Realization

3.1 The Realization of Data Standardization

There are differences between Position Classification Dictionary of People's Republic of China and the performance standard of position classification in the enterprise. The paper designs the correspondence rules for the two standards as below:

- If there is one classification in the enterprise standard which is the same as one classification in the national standard, make a direct conversion from the enterprise one to a national standard one, forming a 1:1 mapping;
- If there is one classification in the enterprise standard which is not same as any classification in the national standard, make a similar conversion from the enterprise one to a national standard one, according to the actual meaning of the classification, forming a 1:1 mapping;
- If there is one classification in the enterprise standard which contents some classifications in the national standard, merge these national classifications together and create a new classification. Then make a conversion from the enterprise one to the new created one, forming a 1:n mapping;
- If there is one classification in the national standard which contents some classifications in the enterprise standard, merge these enterprise classifications together and create a new classification. Then make a conversion from the new created one to the national one, forming a n:1 mapping;
- If there is one classification in the enterprise standard which is not similar to any of the conditions we have mentioned, keep it without any conversion and add it to the performance standard of position classification for the Prediction System as a new classification.

Based on these correspondence rules and Position Classification Dictionary of People's Republic of China, we design a performance standard of position classification for the Prediction System. We convert the twenty-six classifications in the enterprise standard to the fifteen classifications in the system performance standard. The conversion table is shown in Table 1.

Table 1. The performance standard conversion

NO.	The Enterprise Performance Standard	The System Performance Standard
1	Middle-level leaders	The responsible persons
2	General managements	Administrative office staff
3	Financial officer	Economic operators
4	Legal officer	Legal professionals
...

3.3 The Association Rules Model Determination and Visualization

We apply the multi-dimensional improved Apriori algorithm[1] to association rule mining and realize the visualization of mining process and results. C# is the coding language we use to realize the algorithm and visualization. We can freely choose the properties involved in mining and set the minimum support and minimum confidence. The interface of the visualization is shown in Fig. 3.

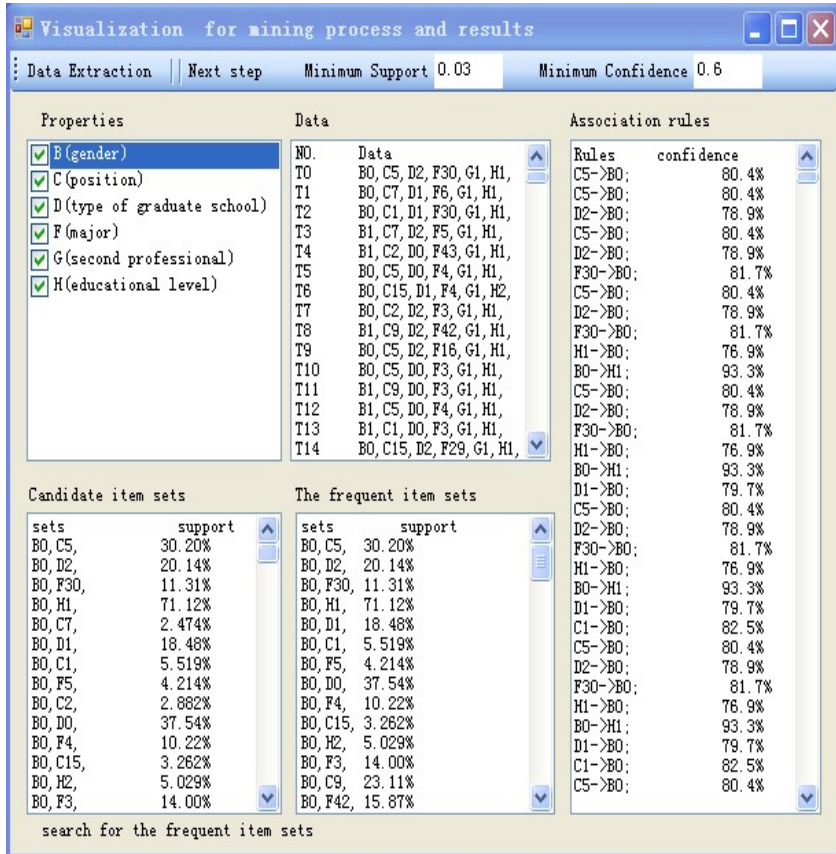


Fig. 3. The interface of visualization for mining process and results

Based on experience, we choose the value of the minimum support as 0.03, and the value of the minimum confidence as 0.6. Association rule mining discovers many association rules. After scientific analysis, we pick up the rules which involved to the five key elements as our interested. The five key elements are gender, type of graduate school, educational level, position and major. The association rules model is shown in Table 4.

Table 4. The association rules model

NO.	Conditions	Results	Support	Confidence
1	The responsible persons	Male	5.52%	82.5%
2	The managements of engineering	Male	30.2%	80.5%
3	The managements of engineering	Full-time undergraduate	35.8%	100%
4	The responsible persons	Full-time undergraduate	5.68%	84.9%
5	Civil engineering	The managements of engineering	4.89%	84.5%
6	Marine engineering	The managements of engineering	5.74%	64.9%
7	985, civil engineering	The managements of engineering	3.62%	90.1%
8	Civil engineering, full-time undergraduate	The managements of engineering	5.76%	84.6%
9	985,full-time undergraduate, civil engineering	The managements of engineering	3.59%	90.3%

4 The Analysis

4.1 The Influence of Single Factor

1. In the machinery manufacturing industry, men with a full-time undergraduate degree often take the managerial position.

In the machinery manufacturing industry, we have 82.5% confidence degree for the rule that the responsible persons are male. We have 80.5% confidence degree for the rule that the managements of engineering are male. There is a relatively high rate for a man to be a business leader or a management of engineering in the machinery manufacturing industry. Universities can adjust the male to female ratio of admissions for manufacturing majors to meet the actual needs of the job market.

In the machinery manufacturing industry, we have 84.9% confidence degree for the rule that the responsible persons have full-time undergraduate degree. We have 100% confidence degree for the rule that the managements of engineering have full-time undergraduate degree. Both business leaders and managements of engineering have relatively high educational background.

2. In the machinery manufacturing industry, graduates from civil engineering major and marine engineering major often take the managements of engineering as their position.

In the machinery manufacturing industry, we have 84.5% confidence degree for the rule that graduates from civil engineering major take the managements of engineering as their position. We have 64.9% confidence degree for the rule that graduates from marine engineering major take the managements of engineering as their position.

4.2 The Influence of Factors

In the machinery manufacturing industry, we have 90.3% confidence degree for the rule that civil engineering graduates from the key universities of the 985 Project with full-time undergraduate degree take the managements of engineering as their position. We have 90.1% confidence degree for the rule that civil engineering graduates from the key universities of the 985 Project take this position and 84.6% confidence degree for the rule that civil engineering graduates with full-time undergraduate degree take this position.

As we can see from the rules, there is a high ratio for civil engineering students from the key universities of the 985 Project with full-time undergraduate degree to be managements of engineering in this enterprise. On the one hand, enterprises have a large dependence of graduates who meet the above conditions; On the other hand, students who meet the above conditions should learn more basic knowledge and strengthen their qualities to be the managements of engineering.

5 Conclusion

In this paper, for machinery manufacturing industry, we conduct a useful exploration and research on potential relation between majors and positions and obtain the primary association rules model. With the deepening of research work, this model will be gradually improved and will provide more effective help for the university major setting and employment trend analysis.

References

1. Xiao, B.: A Multi-dimensional association rules algorithm. In: Chongqing Technology and Business University (Natural Science), vol. 22(04), pp. 339–442. Chongqing Technology and Business University Press (2005)
2. Han, J., Kamber, M., Ming, F., Meng, X.: Data Mining Concepts and Techniques. Machinery Industry Press (2007)
3. Sumon, S., Sarawat, A.: Quality data for data mining and data mining for quality data: A constraint base approach in XML. In: International Conference on Future Generation Communication and Networking, pp. 46–49. IEEE Press, Sanya (2008)
4. Jiang, W., Chen, Z.: The view of the major setting for universities from the Needs Education. Modern Education Science 5, 7–10 (2008)
5. Imhoff, C., Galemno, N., Geiger, J.G., Yu, G., Bao, Y.: Mastering Data Warehouse Design Relational and Dimensional Techniques. Machinery Industry Press (2004)

Research on Partner Selection Issue of Manufacturing Enterprise under Supply Chain Environment

Hong Zhang* and Feipeng Guo

¹ No.280, Xuelin Str., Zhejiang Economic & Trade Polytechnic, Xiasha University Town,
Hangzhou, China
{luqibei, guofp}@hotmail.com

Abstract. Complicated index system and unscientific index weights setting can't adapt the goal of supply chain integration for partner selection. To solve these problems, a comprehensive evaluation index system is established in this paper. Secondly, PCA is used to reduce multiple indicators into a few principal components, yet retaining effective evaluation information. Finally, a partner evaluation model based on improved BP neural network which has self-learning function is constructed. The empirical analysis based on a garment enterprise shows that this model is effective and feasible.

Keywords: Supply Chains, Partner Selection, Principal Component Analysis, BP Neural Network.

1 Introduction

Under supply chain environment, suppliers and manufacturers would not be simple competition relationship. More importantly, they are wanted to build a partnership in order to share information and profits in a period of time. Compared to the traditional supplier selection method, partner selection in supply chain not only concern on cooperation cost, but also considers product quality, flexibility, innovation and others. Therefore, it is more complex and difficult. As a new mode of inter-enterprise relationship management, seamless cooperation is the foundation and key to its success. Thus, how to evaluate and select right partners has become a critical issue.

At present, researches of partner selection include evaluation index system and method. The evaluation indicators were divided into seven categories including 101 indicators by Huang and Keskar [1] which described in detail according to product type and level of cooperation. This is one of newest and systematic outcomes currently. Domestic scholars also did many studies from different views such as uncertainty and efficiency in partner selection. They also have proposed many new methods [2-4]. Overall, methods of partner evaluation include qualitative methods and quantitative methods. The former includes Bidding Law, negotiation method, and intuitive judgments method. The latter includes procurement cost comparison method, AHP, linear weight method, data envelopment analysis, fuzzy comprehensive analysis and artificial neural network algorithms, etc. [5-7].

* Corresponding author.

However, by analyses of relevant literature, most researchers only emphasize on part factors, ignoring other importance factors such as condition of environmental protection, level of application of information technology and collaborative of cooperation. In the weight determination and evaluation value computation, they considered objective fuzziness not enough. Therefore, these methods were often lack of adaptation and scientific. To solve these problems, considering the status and development of partners, a partner selection model based on principal component analysis and improved BP neural network is proposed which can solve the random on indicators selection and subjectivity on weights set.

2 Establishment of Partner Selection Index System

Partner selection is multi-dimensional decision problem. Based on previous studies [1,8], considering the current supply chain management environment, the impact of specific factors for partner selection, according to principle of comprehensive, concise, scientific and reasonable, a index system is proposed including character of enterprise, staff size, reasonableness of management team, credit rating, reasonableness of capital structure, profit situation, cash flow situation, financial soundness, learning capacity of staff, pass-rate of device, service, market share, ratio of total cost of logistics costs, reasonable degree of price, pass-rate of supply, timeliness rate of supply, agility of supply, operation of device, integration of information system, compatibility of strategic concept and management conception, compatibility of management level, compatibility of enterprise culture, compatibility of information system, certifications for environmental quality system, cleaner production level, hardware level and software level of information system, integration of information system, use and maintenance level of information system, mutual ability of manufacturing information, rapid response capability of R&D, funds input of R&D.

3 Partner Selection Model Based on PCA and Improved BP Neural Network

3.1 Principal Component Analysis (PCA)

There are many factors affected partner selection. Different market, different industries, and different forms of cooperation have different emphasis point. Different scholar have different research angle and will get different conclusions which apply to particular scene and has many limitations. Therefore, comprehensive multi-dimensional indicators are established. However, there is a certain correlation relationship among these indicators inevitably. It will waste much manpower if selecting by people's experience. If all indicators are input into neural network, it would increase time complexity and space complexity of network. Finally it would reduce network performance and influence the quality of partner selection. Therefore, principal component analysis method is used to preprocess the indicators.

PCA is a common feature extraction method. It uses dimension reduction to make a number of relevant variables into a few unrelated comprehensive independent variables. These variables contain the principal information of original variables [9-10]. The basic idea is converting k -dimensional random variable $X = (x_1, \dots, x_k)^T$ into a new random vector $Y = (y_1, \dots, y_k)^T$ through space rotation. The new random vector is a linear combination of original variables. There is no correlation among components. This method selects principal components on certain criteria. It calculates the most important principal components, discarding some minor principal components. In that case, PCA can map original high-dimensional data to a new low-dimensional space, so as to achieve the effect of reducing dimension.

3.2 BP Neural Network

BP (Back Propagation) is a multilayered feed-forward neural network based on error back propagation [11-12]. The model is a large-scale parallel processing nonlinear analog system of self-learning, self-organizing, adaptive, whose topology includes input layer, hidden layer and output layer showed on Figure 1.

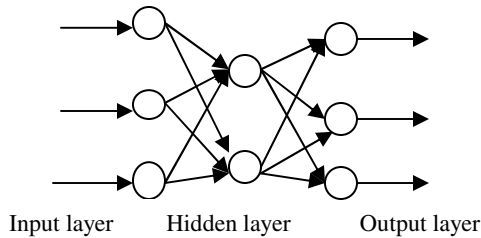


Fig. 1. Three-layer topology of BP neural network

BP algorithm processes a set of training samples iteratively, compares and learns each sample's prediction value with actual class label. Among each training sample, BP modifies the parameters to revise mean square error constantly between network prediction and actual classes, so that the error function declines along negative gradient direction, and achieves convergence ultimately, then learning process stops.

Because the convergence rate of traditional BP neural network is relatively slow when modifying the weight. It often has oscillation phenomenon in training process, and easily falls into local minimum. Therefore, an improved BP neural network algorithm is used which adapt momentum term and adaptive adjustment learning rate [13] to accelerate the convergence process and raise the learning speed.

3.3 Partner Selection Model Based on PCA and Improved BP

Improved BP neural network has strong adaptability, learning and parallel reasoning ability, so it is applied in the paper. Taking account of many indicators of partner selection and some of them are linear correlation, PCA is firstly used to effectively reduce data dimension on the premise of retaining evaluation information. Secondly, a

few principal components are used as the input of neural network. Finally, BP neural network implements selection of cooperation partner by self-learning function.

1) PCA algorithm

Stage 1: Suppose the selection indicators sample matrix of original partners is X .

$$X = (X_1, X_2, \dots, X_p) = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n1} & \dots & x_{np} \end{bmatrix} \tag{1}$$

In above equation, n is number of partners, p is sample dimension.

Stage 2: Execute a standardized transformation of the sample matrix, and convert each index value X_i ($1 \leq i \leq p$) to standardized index Z_i ($1 \leq i \leq p$) according to equation 2 where \bar{x}_j and s_j are mean and standard deviation of X_i respectively.

$$Z_{ij} = (x_{ij} - \bar{x}_j) / s_j \quad (1 \leq i \leq n, 1 \leq j \leq p) \tag{2}$$

Stage 3: Calculate correlation coefficients of the standardized matrix.

Stage 4: Calculate eigenvalues and eigenvectors.

Find out eigenvalue λ_i ($1 \leq i \leq p$) from correlation matrix R using equation 3 in descending order. It also calculate corresponding orthogonal unit eigenvectors b_i ($i=1,2,\dots,p$), each contains p components.

$$|\lambda I_p - R| = 0 \tag{3}$$

Stage 5: Calculate p principal components. The i -th contribution rate of principal component is $\lambda_i / \sum_{i=1}^p \lambda_i$ ($1 \leq i \leq p$). The cumulative contribution rate of principal

components from one to m is $\sum_{i=1}^m \lambda_i / \sum_{i=1}^p \lambda_i$ ($1 \leq i \leq p$).

When the cumulative contribution of eigenvalues is more than 85%, the algorithm stops and obtains m principal components that m is less than p .

2) Improved BP neural network algorithm

Stage 1: Weights initialization.

Stage 2: Calculate net input and output of each hidden layer and output layer. Given the unit j as a hidden layer or output layer, the net input between i and j is I_j .

$$I_j = \sum_i w_{ij} o_i + \theta_j \quad (4)$$

w_{ij} is the connection weights from unit i of previous layer to unit j . If acting on the net input I_j with Sigmoid activation function, then its output of unit j is O_j .

Stage 3: Calculate deviation of each unit j of output layer and each hidden layer k . Update the connection weights and threshold constantly.

$$\Delta W(t) = -l \frac{\partial E_p}{\partial W} + v \Delta W(t-1) \quad (5)$$

$\Delta W(t)$ is modifier value of connection weights obtained from the- t iteration computation, $\Delta W(t-1)$ obtained from previous iteration calculation, v is momentum factor, l is learning rate, E_p is variance between actual value and expected output.

Stage 4: The algorithm stops when global deviation function E is less than a predetermined value or the number of iterations N is greater than a specified value.

4 Analysis

4.1 Data Preparation

Sixty partners' sample is selected randomly from a clothing Co., Ltd during three years. The data include historical transaction data and information given by partners. Before application, it preprocesses firstly and discretizes continuous attribute value. Use relative scoring method to quantify some relatively abstract and ambiguous decision-making factors such as reasonableness of capital structure. The process is using the strongest company in industry as a benchmark, and other unevaluated partner is to be compared with the strongest. For example, 100 point stand for the enterprise A who has the quickest response ability of R&D, and other candidate partner compares with A to determine appropriate score. Both expert and business people score together. The scoring criteria of partners are set based on companies and experts.

4.2 Principal Components

Matlab software is used to get seven principal components after preprocess. The contribution rate of accumulative variance of seven principal components is more than 85%. The relation between principal component and original indices are shown in Table 1, the eigenvalue and contribution rate are in Table 2.

4.3 Establishment and Training of BP Neural Network Model

Two-thirds samples are selected randomly as training samples. The remaining are as test samples. Then, the improved BP neural network model is established and trained. The number of principal components is as the number of input layer nodes. The number of hidden layer nodes uses "test hash" method to train that it selects 6 nodes at

first according to experience. In that case, if the training time is not reach convergence within the required training time, model stops. Then it gradually increases the number of nodes in the hidden layer for retraining. This article tried four times of hidden nodes with 6, 8, 10, and 11. Finally, the best situation is 7 input nodes, 8 hidden nodes, system deviation is 10^{-4} , initial learning rate lr is 0.7, the number of iterations $N = 10000$, and momentum factor $v = 0.6$. The model calls Matlab neural network toolbox and uses traingda learning method to train sample data. The program learned 928 times when deviation value below to 0.0001, then it stops.

After network training is finished, the rest of the 20 companies are selected as test samples. The simulated results are shown in Table 4. As seen from Table 3, the model has good predictive ability on the sample data. The maximum relative error of prediction is 14.8% and the minimum is 0.31%, the average is 2.99%. Therefore, the model has well predicted accuracy and can achieve an effective evaluation of partners.

Table 1. The relation between principal component and original indices

ID	Principal Components	Key features
1	Operating status	Credit rating, Reasonableness of capital structure, Profit situation, Cash flow situation, Financial soundness, Operation of device
2	Sustainability	Learning capacity of Staff, Rapid response capability of R&D, The Funds input of R&D, Certifications for environmental Quality System, Cleaner Production Level, The current of the environment efficiency
3	Products competitive ability	Pass-rate of device, Service, Market share Reasonable degree of price, Ratio of the total cost of logistics costs
4	Supply capacity	Agility of supply, Timeliness rate of supply, pass-rate of supply
5	Cooperation between enterprises	Compatibility of Strategic concept and management conception, Compatibility of Management level, Compatibility of enterprise culture, Compatibility of Information system
6	Level of information	Hardware level and software level of information systems, Integration of Information systems mutual ability of Manufacturing information, Use and maintenance level of Information systems
7	Basic conditions of enterprises	Character of the enterprise, Staff size , Reasonableness of the management team

Table 2. The contribution rate of principal components

Serial number	Eigenvalue	Contribution rate (%)
The first component	52.218	0.2279
The second component	42.926	0.1723
The third component	35.672	0.1412
The fourth component	29.319	0.1102
The fifth component	20.136	0.0792
The sixth component	19.012	0.0735
The seventh component	16.941	0.0702

4.4 Comparison

In order to verify validity of the model, the paper compared with the standard BP neural network algorithm, the improved BP neural network algorithm. The parameters setting are same as this article. The results are shown in Table 4.

Table 3. Deviation analysis of the network training

ID	Actual value	Predictive value	Deviation	Relative deviation (%)
1	75.2	75.6	0.4	0.54
2	67.8	68.1	0.3	0.44
3	30.7	26.8	-3.9	0.14
4	46.6	49.5	2.9	12.7
5	74.8	76.5	0.7	0.94
6	91.3	90.5	-0.8	0.88
7	64.5	65.6	0.9	1.40
8	87.4	86.3	-1.1	1.26
9	56.2	47.9	-8.3	14.8
10	65.4	64.8	-0.6	0.92
11	76.9	78.6	1.7	2.21
12	89.1	90.2	1.1	1.23
13	65.8	66.2	0.4	0.60
14	87.9	86.6	-1.3	1.45
15	57.1	50.2	-6.9	12.10
16	65.4	65.6	0.2	0.31
17	76.5	78.6	2.1	2.75
18	72.9	75.4	2.5	3.40
19	67.9	68.8	0.9	1.33
20	50.7	50.5	-0.2	0.40
Average	68.5	68.2	-9.1	2.99

Table 4. Algorithms comparison

Model	Relative deviation	Iteration times
standard BP neural network	10.23	4892
improved BP neural network	2.76	1097
improved BP neural network(PCA)	2.99	928

It can conclude from Table 4, the accuracy of this model is more than standard BP neural network and less than improved BP neural network. There has little bias by using PCA to reduce data dimension which results in slightly lower detection accuracy. However, this model reaches to the range of deviation in 928 iteration times and avoids resulting into a minimum point.

5 Conclusion

The paper established partner selection model of manufacturing enterprise under supply chain environment and build multi-dimensional comprehensive evaluation index system. Model adopted PCA method to solve randomness in the process of indices selection and avoid personnel subjectivity of indices weights evaluation. Secondly, it used improved BP neural network to evaluate partners.

The application in real clothing enterprise proves this model is feasible. The method not only reduces the model's training time, but also ensures objective and scientific evaluation results. In addition, the model has better generalization ability in other industries and has an application value on similar fields.

Acknowledgments. This work is supported by the Zhejiang Technology Project under grant NO. 2007C31007 and Zhejiang Supply and Marketing Cooperative Research Project under grant NO. 11SS06.

References

1. Huang, S.H., Keskar, H.: Comprehensive and Configurable Metrics for Supplier Selection. *International Journal of Production Economics*, 510–523 (2007)
2. Yun, L., Ye, G., Wei, G.-y.: Dynamic Alliance Partner Selection in the Top-k Problem Optimization Algorithm. *Computer Integrated Manufacturing Systems*, 650–657 (2010)
3. Mao, G.-p., Chao, W.: Evaluation Model for Remanufacturing Products Based on Analytic Network Process. *Computer Integrated Manufacturing Systems*, 2341–2345 (2008)
4. Chen, W.-d., Fan, Z.-p., Yuan, C., Wang, J.-y.: A Supplier Selection Method Based on Fuzzy Group Decision-making. In: *Proceedings of 2005 Chinese Control and Decision Conference (II)*, Haerbin, pp. 623–1627 (2005)
5. Zheng, P., Lai, K.K.: Research on Supply Chain Dynamic Balanced Scorecard Based on Fuzzy Evaluation and Markov Forecast Techniques. *Systems Engineering Theory and Practice*, 57–64 (2008)
6. Jie, C., Yiwen, Y., Senfa, C.: AHP-based Synthetic Evaluation on Performance of Alliance Enterprises Process Management. *Computer Integrated Manufacturing Systems*, 1652–1657 (2008)
7. Basnet, C., Leung, J.M.Y.: Inventory Lot-sizing with Supplier Selection. *Computers and Operations Research*, 1–14 (2005)
8. Chen, Q.: *The Research and Application of the CO-partners Evaluation and Selection in the Environment of Regional CO-manufacturing*. Zhejiang Gongshang University (2009)
9. Babaoğlu, I., Findik, O., Bayrak, M.: Effects of Principle Component Analysis on Assessment of Coronary Artery Diseases Using Support Vector Machine. *Expert Systems with Applications*, 2182–2185 (2009)
10. Howley, T., Madden, M.G., O'Connell, M.L., Ryder, A.G.: The Effect of Principal Component Analysis on Machine Learning Accuracy with High-dimensional Spectral Data. *Knowledge Based Systems*, 363–370 (2006)

11. Ge, z.-u., Sun, Z.-q.: *Neural Network Theory and Matlab R2007 Implementation*. Electronic Industry Press (2007)
12. Liu, Q., Huang, D.: Modeling of Supply Chain Vendor Selecting System Based on AHP-BP Algorithm and Its Application. *Journal of East China University of Science and Technology*, 108–115 (2007)
13. Liu, W.-N., Wang, P., Sun, D.-H., Xie, J.: Forecasting Model of Road Traffic Accident Based on Improved BP Neural Network. *Computer System Application*, 177–181 (2010)

Formal Concept Analysis Based on Rough Set Theory and a Construction Algorithm of Rough Concept Lattice

Haifeng Yang

School of Computer Science and Technology, TaiYuan University of Science and Technology, TaiYuan, People's Republic of China
yhftxy7537750@163.com

Abstract. FCA(Formal Concept Analysis), which is accurate and complete in knowledge representation, is an effective tool for data analysis and knowledge discovery. A new lattice structure named RCL (Rough Concept Lattice) is presented. Using the approximation method of rough sets, we described the extent as approximation extent so that it can deal with uncertainty knowledge. In the end, a construction algorithm CARCL is provided based on it.

Keywords: Formal Concept Analysis, Rough Set, Intent, Approximation, Extent.

1 Introduction

Formal concept analysis, which was presented by Wille R in 1982, is an effective tool for data analysis and knowledge discovery [1]. Every node is a formal concept, which was composed of two elements: Intent and Extent. Intent is description of concept, and extent is objects set, the elements of which have all the attributes of intent. The process of Construction is actually concept clustering. In addition, the relation of specialization and generalization can be shown through Hash map vividly and sententiously. It can be used in information searching, digital library, knowledge discovery, etc [2]. Now, many researches are centered in construction algorithm of concept lattice and its improvement [3] [4], and extension of concept lattice using other theories, such as rough set, fuzzy theory, etc[5-7].

Concept lattice is accurate and complete in knowledge representation, and intent of node describes the common attribute of objects. For an Extent of normal concept lattice, the relation among the attributes of the intent is "and". So, if an object has no only an attribute value of the intent, it has to be got out of the extent set. It is an advantage of FCA, but it is limited in some areas too. If an object has one, another, or some of attributes of the intent, this uncertain data can't be shown, or probably hard to be shown, in FCA. Further more, we have to traverse all the nodes of FCA to get this uncertain knowledge, which will pay out huge cost. For example, a people may have one, another or some of symptoms of one certain illness, but a doctor must not debar this people from the illness just because of the information being uncertainty. So, expressing this uncertain knowledge in concept lattice will be valuable.

First, a new lattice structure, named RCL, is presented based on decision context in this paper. Next, using the approximation method of rough sets, it describes the extent afresh. Last, a Construction Algorithm of Rough Concept Lattice (CARCL) is provided based on it.

2 Rough Concept Lattice

Definition 1. For two random sets A, B, the CADAP (Condition Attribute –Decision Attribute Product) of set A and B, is defined by:

$$A * B = \{(a, b) \mid a \in P(A), b = \langle b_1 b_2 b_3 \dots b_n \rangle, b_i \in B (i \in [1, n])\} \tag{1}$$

Definition 2. Suppose a triple $S = \{O, C \cup D, V_1 \cup V_2\}$ is a decision table where $O = \{x_1, x_2, \dots, x_n\}$, $C = \{c_1, c_2, \dots, c_m\}$, $D = \{d_1, d_2, \dots, d_r\}$, V_1 and V_2 are finite and nonempty sets, the elements of which are called objects, condition attributes, decision attributes, condition attribute values and decision attribute values, and $C \cap D = \emptyset$. The relationships of objects and attributes are described by a binary relation R between O and $C * D$, which is a subset of Cartesian product $O \times C * D$. For a pair of elements $x \in O$ and $a \in C * D$, if $(x, a) \in R$, we say that the object x has the attribute a, or the attribute a is possessed by the object x. So the triple $K_s = (O, C * D, R)$ is called Decision Context which was educed from the decision table S.

Table I is a decision context.

Table 1. A simple decision context

$C \cup D \backslash O$	1	2	3
Fever(c_1)	√	√	√
Cough(c_2)		√	√
Headache(c_3)	√		√
Lack of Power(c_4)	√		
In appetence (c_5)		√	
Conclusion(D)	Grippe (d_1)	Chicken pox (d_2)	Measles (d_3)

Let $K_s = (O, C * D, R)$ be a decision context. Then there is only one partial ordering set respond to it and this partial ordering set can create a structure of normal concept lattice.

Definition 3. For a decision context $K_s = (O, C * D, R)$, operator f and g can be defined by:

$$\forall x \in O, f(x) = \{y \mid \forall y \in C * D, xRy\} \tag{2}$$

$$\forall y \in C * D, g(y) = \{x \mid \forall x \in O, xRy\} \tag{3}$$

Which is to say that f is a mapping from object x to all attributes possessed by x, and g is a mapping from attribute y and all objects which has y.

Definition 4 Let $K_s = (O, C * D, R)$ be a decision context. For an attribute set $Y \subseteq C * D$, we can define the set M and N:

$$M = \{x \mid \forall x \in O, f(x) \cap Y \neq \emptyset\} \tag{4}$$

$$N = \{x \mid \forall x \in O, f(x) \cap Y = Y\} \tag{5}$$

A triple $H (M, N, Y)$ can create any node of concept lattice L , and L is called RCL (Rough Concept Lattice) educed from the decision context Ks .

For a node $H (M, N, Y)$, the element Y is called Intent, which is description of concept; M is called Upper Approximation Extent; and N is called Lower Approximation Extent. The elements of M are objects which have one, another, or some of attributes of Y probably; and the elements of N are objects which have all the attributes of Y certainly.

Definition 5. Let an ordering set $H (M, N, Y)$ be a node of RCL. It is complete on relation R , and it has the properties:

$$M = \{x \in O \mid \exists y \in Y, xRy\}, N = \{x \in O \mid \forall y \in Y, xRy\} \tag{6}$$

$$Y = \{y \in C^*D \mid \forall x \in M, \exists y \in C^*D, xRy\} \tag{7}$$

Definition 6. For two rough concepts $(M1, N1, Y1)$ and $(M2, N2, Y2)$, $(M1, N1, Y1)$ is a sub-concept of $(M2, N2, Y2)$, written $(M1, N1, Y1) \leq (M2, N2, Y2)$, and $(M2, N2, Y2)$ is a super-concept of $(M1, N1, Y1)$, if and only if $Y1 \subset Y2$. If there is no concept (M, N, Y) that is a sub-concept of $(M2, N2, Y2)$ and is a super-concept of $(M1, N1, Y1)$, $(M1, N1, Y1)$ is an immediate sub-concept of $(M2, N2, Y2)$, and $(M2, N2, Y2)$ is an immediate super-concept of $(M1, N1, Y1)$.

Obviously, if $(M1, N1, Y1)$ is a sub-concept of $(M2, N2, Y2)$, $M1$ is the subset of $M2$, and $N2$ is a subset of $N1$.

Definition 7. For two nodes of RCL: r and t . If r is the super-concept of all other nodes of the RCL, we call r the Root-Node; and if t is the sub-concept of all other nodes of the RCL, we call t the Twig-Node.

3 A Construction Algorithm

3.1 The Main Steps of Construction Method

There are three steps in constructing a RCL. First, to get all intents from decision context, we will turn different decision attribute values into one string, and get the ordering sets, the elements of which is one element of condition attributes' power set and a decision string; Next, to get all Upper Approximation Extents and Lower Approximation Extents, we will traverse all records of decision context. M is consisted of all objects possessed one or some of attributes of an appointed intent, and N is consisted of all objects possessed all attributes of an appointed intent. The process of construction is to find immediate sub-concept nodes and immediate super-concept nodes for every new node. We can suppose the node as sextet $(M, N, Y, Parent, Children, no)$.

3.2 The Description of CARCL (the Construction Algorithm of Rough Concept lattice)

INPUT: Decision Table S (O , $C \cup D$, $V_1 \cup V_2$)

OUTPUT: Rough Concept Lattice L

Function1: Get-Intent(S)

```

1 BEGIN
2   FOR every record in S
3     {FOR every v in current record {d=d & v}
4       Des←d}
5   Get power set of condition attributes set C: P(C)
except  $\Phi$ 
6   FOR every d in Des
7     {Create Node h and set h->M, h->N and h->Y  $\Phi$ 
8       FOR every element P in P(C) {h->Y =P  $\cup$ {d};
H←h}}
9 ENDS

```

Function2: Get-M-N(S, H)

```

10 BEGIN
11   FOR every h in H
12     {FOR every Record in S
13       {IF the object possessed all the attributes
of h->Y THEN {h->N←o}
14       FOR every element d in attributes set of
object 'o'
15         {IF d∈h->Y THEN {h->M←o, EXIT FOR}}}
16   FOR every node h in H {IF h->M =  $\Phi$  THEN delete
this node}
17 ENDS

```

Function3: Build-RL(S, H)

```

18 BEGIN
19   L={{(O• $\Phi$ , CUDes,  $\Phi$ , {1}, 0), ( $\Phi$ ,  $\Phi$ ,  $\Phi$ , {0},  $\Phi$ , 1)}
20   FOR every h in H
21     {h->no=Count
22     FOR every node n in L
23       IF h is an immediate sub-concept of n
24       THEN {h->Parent ←n->no; n->Children←h->no;
25         FOR every immediate sub-concept node n'
of n
26           {IF n'->Y  $\subseteq$  h->Y
27             THEN {h->Children←n'->no; n'-
>Parent←h->no;
28               delete n'->no from n-
>Children}}}}
29   Count++; L←h}}
30 ENDS

```

3.3 Algorithm Analysis and Example

We can turn multi-decision-attributes values into one string, get power sets of condition attributes, and initialize all nodes with intents, in Function1. Its time complexity is $O(2^n \times n \times n)$. Upper Approximation Extent M and Lower Approximation Extent N for every node will be initialized in Function2. It's time complexity is $O(n \times 2^n \times |h.Y|)$. And we can build RCL in Function3. It is complete, but inefficient.

Figure1 is a part of RCL based on a decision context which is expressed through table1. #1 and #2 are Root-Node and Twig-Node respectively. For M, the relation among the intent of other nodes is “or”, and for N, the relation among the intent is “and”. For example, M of #9 is {1, 2, 3}, which express that its objects may have fever or cough and may have Chicken pox possibly, N of #9 is {2}, which express that its objects have fever and cough and have Chicken pox certainly.

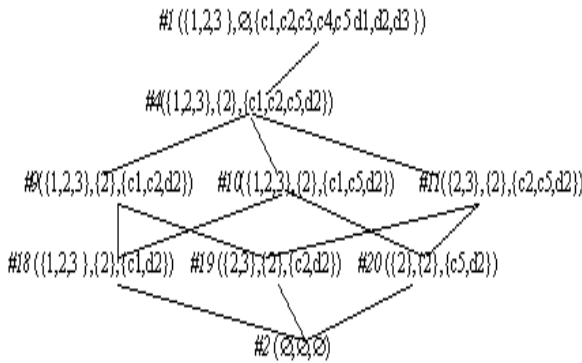


Fig. 1. An Example of a Rough Concept Lattice

4 Conclusions

In this paper, we described extent as approximation extent by using upper (lower) approximation method of Rough Set, so that it can deal with uncertainty knowledge. Based on a decision context, the correctness of CARCL is verified. Next research will be extraction based on RCL and the character and application of RCL.

References

1. Wille, R.: Restructuring lattice theory: an approach based on hierarchies of concepts. M. Rival I Ordered Sets, 415–470 (1982)
2. Hu, K., Lu, Y., Shi, C.: Advances in concept lattice and its application. Tsinghua Univ. (Sci. & Tech.) 40(9), 77–81 (2000)
3. Godin, R.: Incremental concept formation algorithm based on Galois(concept) lattices. Computational Intelligence 11(2), 246–267 (1995)

4. Xie, Z.-P., Liu, Z.-T.: A Fast Incremental Algorithm for Building Concept Lattice. *J. Chinese J.Computers* 25(5), 490–496 (2002)
5. Qu, K.-S., Zhai, Y.-H., Liang, J.-Y., Li, D.-Y.: Representation and Extension of Rough Set Theory Based on Formal Concept Analysis. *Journal of Software* 18(9), 2174–2182 (2007)
6. Qiang, Y., Liu, Z.-t., Lin, W., Shi, B.-s., Li, Y.: Research on Fuzzy Concept Lattice in Knowledge Discovery and a Construction Algorithm. *Acta Electronica Sinica* 33(2), 350–353 (2005)
7. Qiu, W.-g., Zhang, G.-d.: Formal Concept Lattice Based on Composition Rough Operator. *Computer Engineering* 34(24), 16–18 (2008)

Command-to-Submarine Decision Model Research of Formation's Cooperative Anti-submarine Warfare

Jinping Wu, Yi Hu, and Rugang Song

Science and Research Department, Navy Submarine Academy,
Qingdao, Shandong, China
wj_p_wzh@163.com, huyiemail@yahoo.com.cn,
SongRugang1974@sohu.com.cn

Abstract. Synthesized anti-submarine warfare (ASW) capability of submarine and formation in war can be improved through real and effective command-to-submarine in the field of formation's cooperative ASW. As information is the primary content of communication which is the main way of command-to-submarine, the minimum share information set concept of formation's cooperative ASW was presented including its definition and determining the time and content of information exchange. On the basis of fuzzy rules of command-to-submarine decision and ANFIS adopted, the command-to-submarine decision model was built. The corresponding conclusions were obtained according to decision training examples learned and decision verification examples validated.

Keywords: Formation, cooperative anti-submarine warfare (ASW), command-to-submarine decision, ANFIS.

1 Introduction

Synthesized anti-submarine warfare (ASW) capability of submarine who takes part in the formation's engagement can be improved through real and effective command-to-submarine which can make submarine achieve further range and more control sea area in the field of formation's cooperative ASW. The formation's capability of command-to-submarine takes important influence to its ASW action. From the formation's ASW point of view, under the support of command-to-submarine system, submarines and other forces in the formation can easier achieve tactical cooperation, submarine's attack capability and survival probability can be improved. At the same time, the formation's synthesized ASW capability can be improved.

Information is the primary content of communication which is the main way of command-to-submarine. The Information which can be obtained by formation includes battlefield sea area information (electromagnetism information, acoustics environment and atmosphere environment, etc.), all kinds of force entity (enemies, ourselves, friends; under the sea, on the sea, above the sea) state information, weapon and equipment information, action information and ASW decision information, etc. The formation makes command-to-submarine decision which is one important part in formation's cooperative ASW decision in accordance to the information mentioned above.

2 Command-to-Submarine Information Content of Formation's Cooperative ASW

All of the information which can be obtained by formation makes up of the formation's cooperative ASW information set. All of the anti-submarine forces should share common information and make ASW decision in further under the ideal conditions. But there is no need to do like this under the natural conditions mainly because of two aspects: Firstly, the technology limits are faced as a great deal of communication resources will be used (even wasted) because of mass information's high-speed, abundant-capacity and real-time exchange among anti-submarine forces. So the spending for information process is increased which makes the formation's anti-submarine system react slowly even waste engagement opportunity. Secondly, that needs to meet the tactical requirements. The ASW information requirements are different Because of the difference of ASW characteristics among anti-submarine forces. Too much information can create redundancy problems, so achievement of such information maybe takes disadvantageous effects to the commander's decision. Certainly, the information exchanged can't too little so as to that the commander is short of the necessary awareness of battlefield status and lacks further ASW action's accordance.

2.1 The Definition of Cooperative ASW Minimum Share Information Set

The formation's cooperative ASW minimum share information set (MSIS) is the subset of the formation's cooperative ASW information set. The information in this MSIS is chosen from the appropriate information which is come into being from cooperative ASW course. The information can make the anti-submarine forces achieve the same image of the sea state and battlefield situation. When the information is exchanged to the anti-submarine forces taken part in the ASW, the forces are rightly and exactly aware of the sea state and battlefield situation so as to achieve battlefield apperception's cooperation and further make correct tactical decision which ensures better ASW effectiveness.

2.2 Settle of Cooperative ASW MSIS

Two important information exchange aspects should be considered in the settle of cooperative ASW MSIS. First is the opportunity of information exchange (When to exchange), second is the content of information exchange (What to exchange).

Time of Information Exchange. The information exchange during the ASW course has the characteristics that the anti-submarine forces don't exchange information in every time, but exchange information when the sea state or battlefield situation has important changes. So the accident method[1][2] is adopted to settle the cooperative ASW MSIS. The accident means an important change of sea state or battlefield situation such that some force has successfully tracked enemy's submarine, enemy's submarine has reached the attack range of some force's torpedo and enemy's submarine has launched torpedo, etc. during the ASW.

Content of Information Exchange. As a subset, the cooperative ASW MSIS picks up key information from the cooperative ASW information set including:

Index 1. Attributes identification information S_1 of the target including attribute friend or foe, style and shipboard number (or serial number) of the target.

Index 2. Movement information S_2 of the target including bearing, range, course, speed, depth (or height), change ratio of bearing and range as well as the pitch.

Index 3. Threat information S_3 of the target including target threat rank division and threat importance queen.

Index 4. Weapons launched information S_4 such as line-guided torpedo position and guidance information.

Index 5. Tactical decision information S_5 including combat task and order of inshore and formation command position as well as attack and defense decision of anti-submarine forces.

Index 6. Combat stage information S_6 of forces friend including searching, tracking, taking the position, attacking and withdraw maneuvering, etc.

Index 7. Engagement result information S_7 viz. survival information (or damaged information) friend or foe.

According to the information above, if some information deal with by the accident method was signed as $E(S)$, the cooperative ASW MSIS can be denoted as

$$I_{ASW} = \bigcup_{i=1}^7 E(S_i) \tag{1}$$

3 Command-to-Submarine Decision Rules and Fuzzy Expression

ASW information is the basis of the formation's command-to-submarine. Because of some aspects such as the ways of information obtained and the precision of the detection equipment, etc. the information is sometimes fuzzy and non-confirmative. So the rules of command-to-submarine decision should be expressed in the way of fuzzy style.

3.1 Typical Rules of Command-to-Submarine Decision

The typical rules of command-to-submarine decision as following:

Rule 1. IF target range is far AND target shipboard angel is middle AND target is underwater AND target threat degree is weak AND anti-ship missile can't attack

AND line-guided torpedo can attack but needs take the position AND self-guided torpedo can't attack THEN line-guided torpedo attacks.

Rule 2. IF target range is middle AND target shipboard angel is small AND target is underwater AND target threat degree is middle AND anti-ship missile can't attack AND line-guided torpedo can attack immediately AND self-guided torpedo can attack immediately THEN self-guided torpedo attacks.

Rule 3. IF target range is near AND target shipboard angel is small AND target is underwater AND target threat degree is strong AND anti-ship missile can't attack AND line-guided torpedo can attack immediately AND self-guided torpedo can attack immediately THEN torpedo attacks emergently.

Rule 4. IF target range is near AND target shipboard angel is small AND target is underwater AND target threat degree is torpedo-alert AND anti-ship missile can't attack AND line-guided torpedo can attack immediately AND self-guided torpedo can attack immediately THEN submarine defenses the torpedo.

Rule 5. IF target range is near AND target shipboard angel is middle AND target is underwater AND target threat degree is strong AND anti-ship missile can attack immediately AND line-guided torpedo can attack immediately AND self-guided torpedo can't attack THEN submarine defenses enemy's submarine.

3.2 The Fuzzy Expression of the Rules

On the basis mainly of target information from all kinds of sensors or the data results deal with in advance, the formation makes the command-to-submarine decision. Five main factors making influence on the command-to-submarine decision were selected as the fuzzy rules input of the command-to-submarine decision knowledge including:

Target range. Its fuzzy set can be denoted as

$$A_1 = \{far, middle, middle-near, near\}.$$

Target shipboard angel. Its fuzzy set can be denoted as

$$A_2 = \{great, middle, small\}.$$

Attacking probability of the line-guided torpedo. Its fuzzy set can be denoted as

$$A_3 = \{canattack, can'tattack\}.$$

Attacking probability of the self-guided torpedo. Its fuzzy set can be denoted as

$$A_4 = \{canattack, can'tattack\}.$$

Target threat degree. Its fuzzy set can be denoted as

$$A_5 = \{torpedoalert, strong, middle, weak\}.$$

Among the five factors above, the range, shipboard angel and threat degree of the target is a kind of succession knowledge. Attacking probability of the self-guided and line-guided torpedo is discrete knowledge but can be expressed as succession knowledge. The five factors can be denoted by membership function of normal school. For example, the membership function of target range can be denoted as

$$\mu_{Di}(x) = e^{-\left(\frac{x-a_i}{\sigma_i}\right)^2}, i = 1,2,3,4 \tag{2}$$

4 Model Inputs and Simulation Results

The course of command-to-submarine is a typical non-linear fuzzy system whose indexes mainly include target range, target shipboard angle, target type, target threat index and weapon attack probability. All of these indexes can be denoted in the way of fuzzy set.

4.1 The Inputs of the Model

Supposed that there are 5 inputs in the fuzzy system which can be given as

$X = \{x_1, x_2, x_3, x_4, x_5, x_6\} = \{\text{target range, target shipboard angle, line-guided torpedo attack probability, self-guided torpedo attack probability, target threat index}\}.$

Then the fuzzy set of these indexes can be denoted as

Target range. Value (0~150cable). Far(90~150cable), middle(50), middle-near(30~50 cable), near(0~30 cable). The center of the membership function is 150, 70, 40, 0.

Target shipboard angle. Value (heavy, middle, small). heavy(>45°), middle(15~45°), small(<15°). The center of the membership function is 90, 30, 0.

Line-guided torpedo attack probability. Value (0~100second). Can attack(0~50second), can't attack((50~100second). The center of the membership function is 100, 0.

Self-guided torpedo attack probability. Value (0~100second). Can attack(0~50second), can't attack((50~100second). The center of the membership function is 100, 0.

Target threat index. Value (0.0~1.0). Torpedo alert(0~0.1), strong threat(0.1~0.4), middle threat(0.4~0.7), weak threat(0.7~1.0). The center of the membership function is 0, 0.25, 0.55, 1.0.

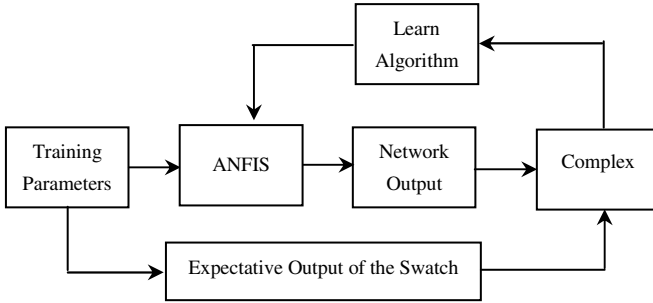


Fig. 1. The Learning Course of Command-to-Submarine Decision Based on ANFIS

The course of the command-to-submarine can be presented in the way of ANFIS (Adaptive Neural Fuzzy Illative System) as shown in Fig.1. The command-to-submarine decision result can be obtained by its output variable as following.

$Y = \{y_1, y_2, y_3, y_4\} = \{1 \text{ line-guided torpedo attack, } 2 \text{ self-guided torpedo attack, } 3 \text{ defense torpedo, } 4 \text{ defense submarine}\}.$

4.2 Simulation Results

The programming simulation was executed by MATLAB neural network toolkit. The learning to the training switch was processed. The structure of the training switch is shown as Table 1.

Table 1. The Structure of Training Switch

ID \ Inputs	1	2	3	4	5	6	7	8	...
Target Range	76	23	42	129	25	43	43	86	...
Target Shipboard Angle	29	12	47	9	8	15	43	31	...
Line-guided Torpedo Can Attack	0	0	100	10	100	0	0	100	...
Self-guided Torpedo Can Attack	16	100	0	28	100	0	0	16	...
Threat Degree	0.8	0.1	0.4	0.9	0.0	0.3	0.4	0.8	...
Decision Result	1	4	2	1	3	2	1	2	...

The change curves of the fuzzy membership function and the step before and after the training were obtained under the precondition of meeting the requirements of precision. For example, the change curves of the target range were shown in Fig.2 and Fig.3.

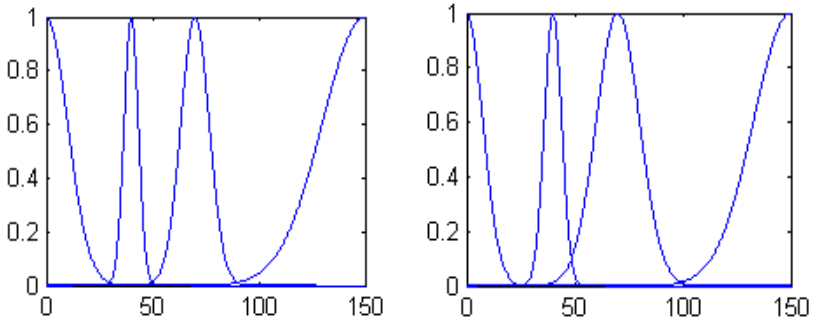


Fig. 2. The Range's Change Curves of the Fuzzy Membership Function

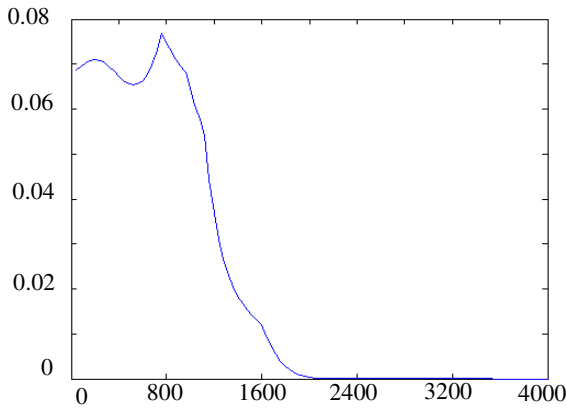


Fig. 3. The Change Curve of the Step-Length

5 Conclusions

The conclusions can be obtained from analyzing the curves of simulation results as following:

Conclusion 1. Through verifying by typical verification swatch, a conclusion can be obtained that the ANFIS model was in commendable accordance with the non-linear and non-confirmative fuzzy relationship among every evaluation index during the course of command-to-submarine decision.

Conclusion 2. The command-to-submarine decision model designed meets rightly the requirements of the formation command-to-submarine decision. The input data roots directly in the regulation database of cooperative ASW and can preferably reflects the basic command-to-submarine course of formation's cooperative ASW.

Conclusion 3. The model raises high demand to the precision of the initial membership function. The convergence of the model depends on the membership function. The requirements of the formation's command-to-submarine decision can be met through adjusting appropriately the mean and mean square deviation of membership function.

Conclusion 4. The speed of convergence can be made more reasonable through making good use of changing step (learning ratio).

Conclusion 5. At the aspect of formation's command-to-submarine decision, the fuzzy neural network isn't easy to run into local minimal value point and has more quickly convergence rate and better evaluation results than traditional neural network based on BP algorithm.

References

1. Teng, K.N., Yang, C.Z.: The Minimum Share Information Set Research of Warship Formation Anti-Air Warfare Based on CEC System. *Journal of Naval Engineering University* 19(4) (2007)
2. Li, L.S., Yu, H.X.: The Minimum Share Information Set of Cooperative Air Warfare. *Fire Command and Control* 26(2) (2001)
3. Akbori, F.: Autonomous-Agent Based Simulation Of Anti-Submarine Warfare Operations With The Goal of Protection A High Value Unit. In: NPS (March 2004)
4. Zhang, H.B., Zhao, J.Y.: The Decision Model Research of Minimum Share Information Based on Rough Set Theory. *Computer Science* 30(5) (2003)
5. Nie, W.D., Kang, F.J.: The Simulation Research of Cooperative Engagement Capability In the Field of Underwater. *Journal of System Simulation* 16(2) (2004)
6. Cheng, Q.Y.: Indefinite Dynamic Conflict Decision Theory and Its Application Research in Military Command. *Journal of Beijing Aviation and Aerospace University* (February 2002)
7. Lin, F.D., Du, Y.P.: The Effectiveness Analysis of Warship Formation Searching Submarine. *Fire Command and Control* 33(2) (2008)
8. Askeland, J.H.: Network Centric Warfare Communication Integrated In Submarine CMS. *Naval Forces* (November 2003)
9. Li, D.F., Xu, T.: Analysis and Application of Naval Engagement. National Defense Industry Press, O.R. Beijing (August 2007)

A Context-Aware Model Based on Mobile Agent

Ying Wang and Xinguang Peng

Colledge of Computer Science and Technology
Taiyuan University of Technology
Taiyuan, China
ablewy@163.com, sxgrant@126.com

Abstract. With the development of Internet120, the information on the Web is more and more abundance. In this paper, combining Mobile Agent theory and Context-aware, a kind of Context-ware model based on Mobile Agent has been brought forward. The function of analyzing actual requirement of users which combines their context information, utilizing Mobile Agent to aware, process and manage context information, and providing precise knowledge to users are implemented. At last a Context-aware prototype system based on Mobile Agent which is on the Aglet software platform is realized, it can aware of user profile context and the experiment preferably educes the characteristic of Context-aware.

Keywords: Context-aware, Mobile Agent, Aglet.

1 Introduction

With the sustainable development of Internet, the knowledge on the Web is growing rapidly. At present, people usually search knowledge by using the method of inputting key words. If the key word is too general, we may get necessary knowledge from a lot of results. Even the keyword is specific enough; it is also difficult for us to obtain actual required knowledge. The key problem is that the system, which provides knowledge searching service didn't aware of users' context information, such as location, profile, time, and so on. Therefore, providing appropriate knowledge searching service through Context-aware for users is an important task, it has profound academic significance and applied value.

In this paper, a Context-aware model has been established. By using the model required precise knowledge which combining with the user's context information is provided. We use mobile Agent to aware, process, control and manage the context information.

2 Context-Aware

The concept of context has been used in many ways in different areas of computer science, such as "context-sensitive help", "contextual search", "multitasking context

switch”, “psychological contextual perception”, and so on. We only focus on the context used by applications in mobile computing¹. The context can be defined as “any information that can be used to characterize the situation of an entity. And an entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves” [1]. The typical context such as profile, location, identity, and time can be used in many applications.

When providing mobile services to service object, context information is a kind of important service basis. The accuracy and credibility of this basis is called quality of the context. The quality of Mobile service is determined by it to a large extent. And it can be examined from four aspects including automation, availability, dynamicness and correctness [2].

The Context-Aware computing is a kind of application which can adapts to discovered context automatically, by changing the application’s behavior. Context-Aware computing has received comprehensive attention from many overseas research institutions since the 1990s, and many related applications had been developed, such as Shopping Assistant [3] from AT&T Bell Laboratories, Office Assistant [4] from MIT Media Laboratory, and Conference Assistant [5] from Future Computing Environments (FCE) at the Georgia Institute of Technology.

3 Mobile Agent Technology

In the early 1990s, the General Magic Company puts forward the concept of Mobile Agent in the Telescript commercial system. In brief, Mobile Agent is a type of software agent, with the feature of autonomy, social ability, learning, and most importantly, mobility. More specifically, a Mobile Agent is a process that can transport its state from one host to another in heterogeneous network, with its data intact, and be capable of performing appropriately in the new environment.

Mobile Agents provide many advantages. First, they reduce network overloading allowing the code to be executed at data location, instead of fetching remote data from code’s emplacement. Also, network latency is decreased. Agents can be used to execute locally where the control is required, reducing latency time in real-time applications. Mobile Agents provide an asynchronous and autonomous execution which is ideal to work with in environments with expensive or fragile network links. Heterogeneity is a natural feature inherent in mobile agents. They are hardware and transport layer independent and hence they provide optimal conditions for uniform system integration. Robustness and fault tolerance are two more advantages easily provided by Mobile Agents. Mobile Agent system consists of mobile Agents and Mobile Agent infrastructure [6]. Mobile Agent infrastructure implements transportation of Agents between hosts based on Agent Transfer Protocol ATP, and offer execution environment and service interface for Agents. Agents are executed in infrastructure. They communicate with each other based on Agent Communication Language ACL and get services from infrastructure.

4 Context-Aware Model Based on Mobile Agent

The appearance of Context-aware theory and Mobile Agent technology provides a new method to traditional knowledge search system. The use of Context-aware and Mobile Agent change the traditional service mode and search method. It can help users acquiring the knowledge related to their real intentions, while reducing the network bandwidth dependent degree of traditional system. Therefore the service capability can be improved greatly.

Context-aware model can analyze the user's actual demand combined with his context information, and provide precise knowledge for users by utilizing Mobile Agent to aware, process, control and manage with the context information. This kind of design greatly facilitates the design and implementation of Context-aware mobile service. How the Mobile Agent used in Context-aware model can be categorized as follows:

- Extract context information from heterogeneous execution environment;
- Comprehensive process the obtained context information to determine the context of mobile client;
- Provide adaptive service to the specific context.

The structure of the Context-aware model based on Mobile Agent is shown in fig.1, as can be seen from the figure, the Context-aware model is divided into three layers: context extraction layer, Context-aware process layer and context submission layer.

The function of context extraction layer is the extraction of different context. It mainly consists of some context extraction Mobile Agents. Various context extraction Mobile Agents are used for extracting different types of context. The mobile schedule is established for Agents to accomplishing context extraction.

Context-aware process layer realized the function of representing context from context extraction Mobile Agents normatively, scheduling context with user record reasonably, and submitting the result to knowledge search system interface. It composes of a control Agent, some context process Agents and a context dispatcher Agent.

The knowledge search system interface in context submission layer implements the acceptance of user context witch is submitted by context dispatcher Agent.

The aware of context get started with monitoring and acquiring original data by context extraction Mobile Agents. Different context extraction Mobile Agents are applied to different types of context. Their function is the management of Context Sensors which is used to extract data. In the model there are three kind of context should be aware: natural context, technical context and social context.

The function of Context process Agent is converting original context related to current application which is extracted from computing environment to context information with clear semantic and unified format. When context information is formed, Context Slot is used to representing it. Its structure is shown in fig.2.

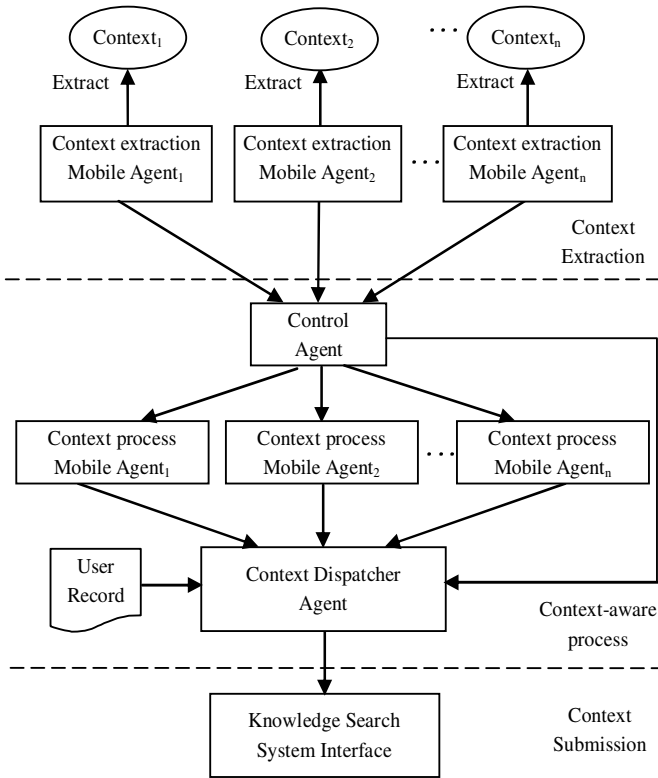


Fig. 1. Context-aware model based on Mobile Agent

PRI	TIME	WAIT	R	CONTENT
-----	------	------	---	---------

Fig. 2. Structure of Context Slot

The meaning of each field is listed as follows:

- PRI: priority of context;
- TIME: maximum duration of response;
- WAIT: maximum waiting time;
- R: a kind of tag;
 - If the value of R is 1, it means this context can interrupt the executing of current context.
- CONTE: the content of context information.

A knowledge base is built to store context information according to the semantic constitution of context information, meanwhile for facilitating dynamic modification

of context information. In order to provide precise search result to users, it is not sufficient for decision making only aware of one kind of context. The system sent out application communication request to higher level when context information is formed. Then reasonable dispatch will occur according to the priority of context.

Compared to traditional search technology, using Mobile Agent technology to Context -aware and WEB search has advantages of utilizing Internet bandwidth effectively, dispensing with persistent network connection and load balancing.

5 Prototype System

Context-aware based on Mobile Agent prototype system: CABMA is a Context-aware prototype system developed on Mobile Agent platform. The function of it is as follows: firstly, users' context are aware of by using Context-aware model when they login for searching knowledge, then necessary knowledge are searched with users' context, and finally search results are returned to users. A Mobile Agent platform Aglet is adopted in CABMA. It is a pure Java development mobile Agent platform of IBM Company.

The structure of CABMA system is shown in fig. 3. The system is made up of four parts: user interface, control subsystem, user Context-aware subsystem and knowledge search subsystem. User Profile context directly influences the habit and interest of users' knowledge search, and is the key part of the Context-aware system. Therefore in the implementation of system, a user Profile Context-aware experiment is conducted on the basis of our equipment conditions. When a user sends out a search request to system, a Context-aware Agent is generated in local. Then this Mobile Agent is sent to remote host to aware of user Profile context. As a result user Profile context can be extracted. Finally, appropriate knowledge according to user Profile context will be searched in local resources and search results can be returned. Fig. 4 shows the principle of experiment. In the experiment, two PCs installed Aglet platform are used, one is server host named "wq", the other is client host named "pmy" which is the destination host of Context-aware .Also a user Profile library is constructed to store user Profile.

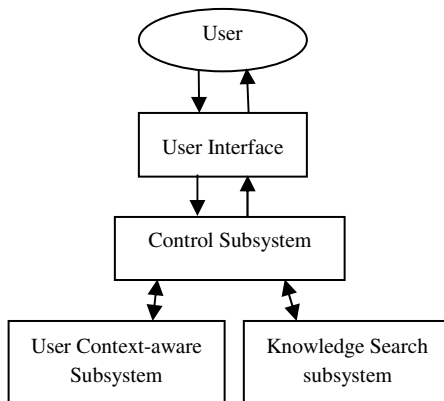


Fig. 3. The structure of CABMA system

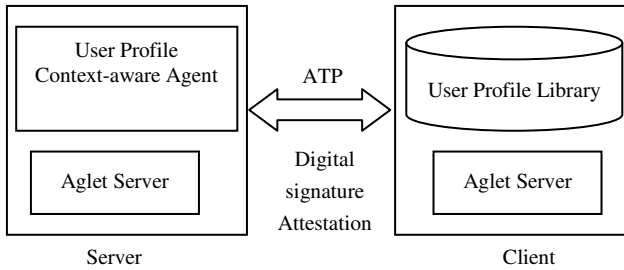


Fig. 4. The principle of CABMA experiment

After installing the Aglet platform and setting environment variables, then we can create our own Mobile Agent based on Aglet platform. The management interface of Aglet server is called Tahiti, and all the management of Aglet are carried through here. Needed Agent will chosen to execute and some operations can be done here, such as: send to and recall from a remote host, destroy an Agent, etc. In the experiment, a Mobile Agent example called "ContextAglet" is created to aware of user's Profile context. The running interface of ContextAglet is shown in fig. 5. The destination host list of Context-aware is placed in Address Book. In our experiment, "pmy" is the destination host and Click "GO" button, a Mobile Agent called "ContextAglet" will be sent out. When it reaches to the host with address "Atp:// pmy: 4434" , user Profile can be extracted from Profile library by "ContextAglet" Mobile Agent, meanwhile search result will be returned. The running status of Mobile Agent is shown in fig. 6.

Thus we can read the basic information of user pmy. They are Name: pmy; Sex: Female; Age: 25; Job: teacher; Hobby: Computer. According to these user context, required knowledge can be provided to users.

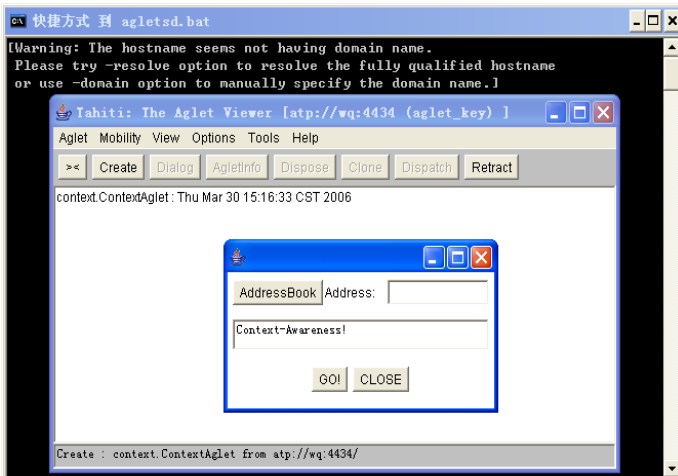
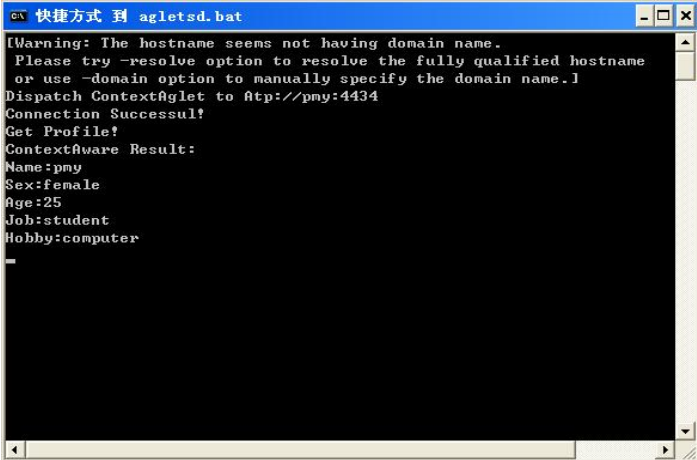


Fig. 5. ContextAglet running interface



```
快捷方式 到 agletsd.bat
[Warning: The hostname seems not having domain name.
Please try -resolve option to resolve the fully qualified hostname
or use -domain option to manually specify the domain name.]
Dispatch ContextAglet to Atp://pny:4434
Connection Successful!
Get Profile!
ContextAware Result:
Name:pny
Sex:female
Age:25
Job:student
Hobby:computer
```

Fig. 6. ContextAglet running status

6 Conclusion

Context-aware is a rising research field. Research in this field is just atarting at present. Development of Context-aware applications require two aspects of support: hardware and software. Compared to some research abroad, our research scope has limited because it is unable to obtain more kinds of sensors. In this paper, the design structure and realizing technology of Context-aware model based on Mobile Agent are systematically introduced. On the basis of it, the design and experiment of CABMA system is conducted. Basic functions of Context-aware syatem based on Mobile Agent are successfully realized. User Profile context are awared in experiment.

Acknowledgment. This paper is Supported by the Natural Science Foundation of Shanxi Province under Grant No. 2009011022-2 and Shanxi Scholarship Council of China Grant No. 2009-28.

References

1. Chen, G.L., Kotz, D.: A Survey of Context-Aware Mobile Computing Research. Dartmouth Computer Science. Technical Report TR2000-381, 2-3 (2001)
2. Schilit, B., Adams, N., Want, R.: Context-aware computing applications. In: IEEE Workshop on Mobile Computing Systems and Applications, pp. 85-90. IEEE Computer Society Press, Los Alamitos (1994)

3. Asthana, A., Cravatts, M., Krzyzanowski, P.: An indoor wireless system for personalized shopping assistance. In: IEEE Workshop on Mobile Computing Systems and Applications, pp. 69–74. IEEE Computer Society Press, Los Alamitos (1994)
4. Yan, H., Selker, T.: Context-aware office assistant. In: The 2000 International Conference on Intelligent User Interfaces, pp. 276–279. ACM Press, New York (2000)
5. Dey, A.K., Futakawa, M., Salber, D., Abowd, G.D.: The Conference Assistant: Combining Context-Awareness with Wearable Computing. In: 3rd International Symposium on Wearable Computers, pp. 21–28. IEEE Computer Society Press, Los Alamitos (1999)
6. Ghezzi, C., Vigna, G.: Mobile Code Paradigms and Technologies: A Case Study. In: Rothermel, K., Popescu-Zeletin, R. (eds.) MA 1997. LNCS, vol. 1219, pp. 39–40. Springer, Heidelberg (1997)

A Mode of Storm Flood Forecasting DSS Establish Ion

Yaoxing Yan and Yutao Cao

College of Water Resources Science and Engineering
Taiyuan University of Technology, TYUT
Taiyuan, china
tyutwater@163.com

Abstract. Storm flood forecasting can not accurately describe using one or several models. The chaos of the storm runoff systems make the prediction model must face with difficulty which can not be solved now. This paper will provide a new method to solve the problem proposed before. Use the information of the historical storm flood, establish rainstorm forecasting DSS knowledge base, take advantage of correlated identification between real-time storm flood and knowledge base, and offer important information about flood forecasting in order to provide expert consultations decision support of real-time storm flood forecasting.

Keywords: Flood forecasting, Artificial intelligence, DSS, Gray correlation.

1 Introduction

Flood forecasting is very important in national economic construction and people's life, and the flood causes huge losses every year. Therefore, the rainfall flood forecasting becomes a major and practical significant research topic. In recent years, flood forecasting technology combines with computer technology, and the flood forecasting level has been greatly improved with computer database technology, artificial intelligence technology and related soft wares development, particularly using Web technology. After 30 years development, flood forecasting target goes through from initial simple automation of the on-line real-time, to the combination of graphics processing technology, the flood forecasting model for the mathematical precision processing technology, and to the computer consultation, human intervention Interactive forecasting system, the idea and design of the flood forecasting has great innovation.

As we know, there are many storm flood forecasting models, which the parameters are very complexity, the algorithms of mathematical theory are very rigorous, also numerical methods are very advanced, but the forecast errors are extremely huge. In fact, flood forecasting is based on the storm forecasting, but the forecast of the rainstorm accuracy is not high today, so the flood forecasting accuracy can imagine. The successful experience in some of our big rivers flood forecasting tells us that, the experts' consultation is a key step, in addition to meteorological data analysis and calculation large number of flood forecasting, and the information of flood forecasting requires expert consultations before publication in our country. This shows that the accuracy of flood forecasting not only entirely depend on the existing flood forecasting model, algorithm, but also experience, ability of fuzzy comprehensive and expert

knowledge. Therefore, it is very important to establish the expert knowledge base and flood forecasting DSS (decision support system).

In this paper, the author designed and developed a storm flood forecasting DSS combine with successful experiences of storm flood forecasting in our country. And this can provide a new way for expert consultations and rapid, accurate, scientific, rational prediction through intelligent processing by computer and full use of historical data information.

2 Mode Analysis

The accuracy of flood forecasting depend on many aspects, including: precipitation (including intensity and volume), rainfall distribution, watershed land surface conditions, surface soil moisture and distribution before rainfall, runoff model selection, calculation methods, and analysis of historical flood, results consultation and forecast information publish ion. First, rainfall is the most basic part of the forecast, China is currently planned for the construction of hydrological forecast site, and with the development of information collection technology, the collection of rainfall data problems will gradually improve and meet the requirements of flood forecasting. Second problem is the soil entropy forecast, with the development of satellite technology and improvement of soil entropy collecting device's wireless transmission technology, these fundamental issues of flood forecasting will be gradually solved.

Flood forecasting is a typical gray system. The flood size may be different when the same storm happened in the same basin in different time, whereas they may have huge similarity or same when different rainfall course in different basin. With the development of national economic, environmental change, construction and other man-made changes in watershed land surface, and the flood forecasting can not be solved only by one model, therefore it can not solve the fundamental problem depending on try to use more and more complex model.

This paper aims to seek a way similar to artificial intelligence method, which making fully use of information of historical and real-time storm flood information, and to establish of flood forecasting DSS for expert consultation. This approach will increasingly reflect the superiority of the scientific and rational along with time and the accumulation of storm flood data. This method can be accomplished by using the computer's artificial intelligence design, expert knowledge systems and large database analysis system. And flood forecasting structure will no longer be a tree structure, but the three-dimensional network structure. In this structure, the prediction will not be easy response treatment, but can provide incident handling processes and results according with the type and magnitude which has happened; and the system is no longer a simple decision support, but can intelligently participate in decision-making.

3 Mode Establish Ion

3.1 Storm Flood DSS Establish Ion Idea

Storm flood DSS establish ion idea is as follows:

1. To establish the basic database of history storm, flood, surface conditions, soil moisture and etc.

2. To establish storm flood forecasting knowledge database under historical data. Which can use the computer to calculate the size of every storm flood, the soil moisture at different time points, etc, and can analysis and calculate the flood course under soil moisture and storm flood caused by responding wet at different month and time points. Then take the results as records of knowledge base of the flood.

3. According with the time period of the storm rain occurred, the real forecast of storm rain, soil humidity and other data, using gray correlation analysis to calculate, and finally giving one or a few records of best correlation.

4. According to characteristics associated with the consolidated results, accompanied by the necessary amendments, such amendments rainfall, soil moisture correction, given by the results.

5. Submit the results as a basis for decision support of expert consultation.

6. According to the analysis and mending results of prediction and the measured, add the results to the knowledge database for decision-making by the next flood.

Storm flood DSS uses gray relational analysis method, mainly due to storm rain flood system itself is a typical gray system. And compare with the course of measured storm flood and history flood, gray relational analysis is most effective. Gray by gray theory correlation analysis, multiple factors are given the highest level of the history associated with the flood as the flood forecast, supported by heavy rainfall period of the fuzzy analysis method, through a number of association analysis, and knowledge of historical floods Contrast, the results provide forecasts for flood forecasting in consultation with the decision support results.

Gray system theory correlation analysis is according to the each various factor curve similarity degree to determine the correlation. Through quantitative analysis of dynamic process of development, the method completes the comparison of statistical data of the system, and obtains the gray correlation between reference and compared series. The bigger between reference and compared series, their relationship is closer, and the direction and speed is more close to compared series. If we take the measured storm flood course as a time series and the same as historical storm flood, then we bring them to carry on gray correlation analysis and determine the most likely flood finally. The following will show an example about gray correlation calculation of the measured rainfall and historical flood.

3.2 Gray Relational Analysis Method Steps

The specific calculation steps of Gray relational analysis method are as follows:

The first step: to determine the number of columns.

Assuming the number of measured storm rainfall as $Y = \{Y(k) | k = 1, 2, \dots, n\}$;

historical storm rainfall as $X_i = \{X_i(k) | k = 1, 2, \dots, n\}$, $i = 1, 2, \dots, m$.

The second step, dimensionless of variables

Because the system daters in each factor may vary in different dimensions, it is not easy to compare and is difficult to get the right conclusion. Therefore, during the gray correlation analysis, the data are generally required to deal with the dimensionless as below.

$$x_i(k) = \frac{X_i(k)}{X_i(k)}, \quad k = 1, 2, \dots, n; \quad i = 0, 1, 2, \dots, m. \tag{1}$$

The third step, to calculate correlation coefficient

The correlation coefficient between $Y(k)$ and $X_i(k)$ is:

$$\xi_i(k) = \frac{\min_i \min_k |y(k) - x_i(k)| + \rho \max_i \max_k |y(k) - x_i(k)|}{|y(k) - x_i(k)| + \rho \max_i \max_k |y(k) - x_i(k)|}. \tag{2}$$

If $\Delta_i(k) = |y(k) - x_i(k)|$, then

$$\xi_i(k) = \frac{\min_i \min_k \Delta_i(k) + \rho \max_i \max_k \Delta_i(k)}{\Delta_i(k) + \rho \max_i \max_k \Delta_i(k)}. \tag{3}$$

$\rho \in (0, \infty)$ is called the resolution factor, when ρ get more smaller and the resolution will get more greater. The general value of ρ interval is $(0, 1)$, and when $\rho \leq 0.5463$ the resolution is the best. Usually $\rho = 0.5$.

The fourth step, calculation of correlation

Because the correlated coefficient is the value to compare series and reference sequence in each moment (the points in the curve), and the data is more than one number, so the information is too scattered for overall comparison. Therefore it is necessary to concentrate all correlated coefficient each moment (the points in the curve) to one value that is to get their average, and take it as value between compared series and reference sequence. The formula of correlation is as follows:

$$r_i = \frac{1}{n} \sum_{k=1}^n \xi_i(k), \quad k = 1, 2, \dots, n \tag{4}$$

The fifth step, the arrangement of correlation

The correlation is ordered by size, and if $r_1 < r_2$, and then the reference sequence y is more similar with compared series x_2 .

Calculate the average of all correlations after compute the series correlation between $y(k)$ and $x_i(k)$, and the average r_i is the correlation.

Soil moisture is not need to calculate correlation, but according to the size of the relative error to judge whether it can be used or not.

Select the best correlation results or through a certain amendments and then submit the decision support results.

The basis work flow of rainstorm and flood forecasting model DSS under this mode is shown in Figure 1.

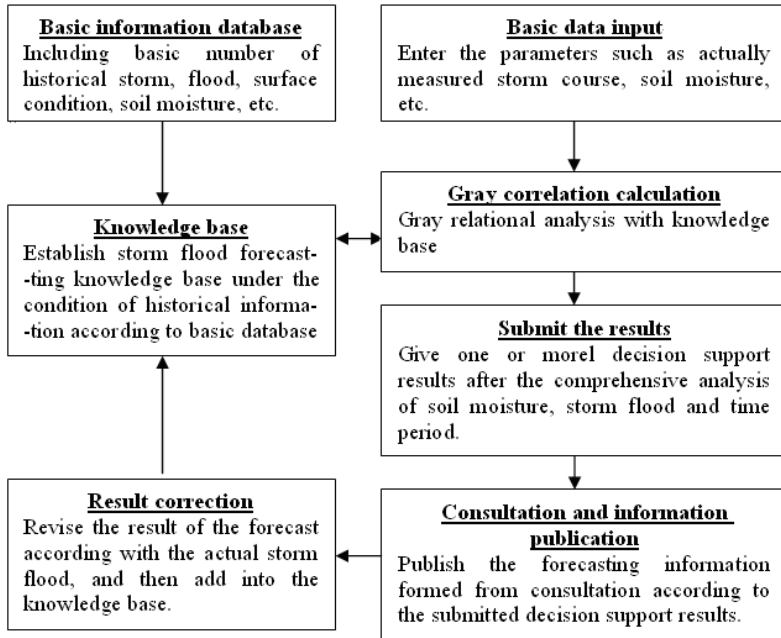


Fig. 1. The basis work flow of rainstorm and flood forecasting model DSS mode

The amendment results are divided into two steps, one is correction of forecasting consultation results, and another is amendment of flood after the storm flood happened. The former amendment is a collection of experiences and theories, such as the linear and nonlinear correction about different storm size, and different linear and nonlinear correction of wet soils. This correction phase mainly amend through consultation. Whereas, the latter correction is simpler, as long as the amendment according with the real measured storm floods and corresponding soil moisture, etc. and then we can directly add them into the knowledge base.

4 Application

Take measured storm flood data from year 1969 to year 2006 at Wenyuhe River hydrological stations in Shanxi as an example. This station did not take place bigger flood since 1969, and we can predict and test only depending on the small and medium storm flood. If the annual maximum flood peak discharge is less than $20\text{m}^3/\text{s}$, and the flood volume is less than 200 million m^3 not be considered, then there only leave 14-year data. Statistic period for each storm flood is 3h and the general storm last no more than 5 periods. The storm lasted 7 periods in the August 15, 1977 and 9 periods in 16 September 1978, and other 9-year storm is all 5 periods. Therefore, using division storm runoff way to cut up the flood course which is exceeded 5 periods, then establish the knowledge base which mainly include soil moisture before the storm (or antecedent rainfall), heavy rainfall, the corresponding flood.

Adopting recent June 14, 2002 storm for forecast verification, this storm has 5 periods, the antecedent rainfall $W=60\text{mm}$ after calculation. Compare with the knowledge base, only the August 15, 1977 antecedent rainfall= 65mm , close with it, and their soil wet relative error is less than 10%, which is can be considered as satisfying the demand. The two storms' correlation is high to 0.91 after calculation, so we may assume them have higher similarity. Considering the two storm can both make the surface soil saturated, that is to say, they can reach the maximum antecedent of the basin $W=80\text{mm}$, so the forecast flood peak and flood volume can simply use storm ratio to linearly revise the result of the prediction.

5 The References Section

Storm flood forecasting is an eternal subject, the flood runoff is a chaotic system, and it can not be described clearly through one or more complex models. The mode established in this paper is an attempt. Because of data conditions, it failed to carry on more storms or other basins for test the correct of the mode, and prediction amendment are not all linear correction, so there are many issues to be further improved.

With the materials of storm flood and soil moisture accumulation, the knowledge base will become gradually enrich; and the mode we put forward will have a better performance. We may use C/S technology to forecast and consultate through the network, release real-time information, and comply remote consultation and information dissemination functions.

References

1. Zou, B., Gao, J., Li, Y.: Development and application of Water Information System-Hydrological Forecasting System. *Yangtze River* 42(6) (2011)
2. Zhao, X.: Research of flood Real-time Forecasting model. *Chinese Agricultural Science Bulletin* 26(13) (2010)
3. Wu, Y., Wang, L.: Analysis on storm-flood fo Minjiang River. *Forecast Practice, Yangtze River* (July 18, 2010)
4. Wang, C.-p., Wang, J.-s., Liang, T.-h.: Application of artificial intelligence in The flood forecasting. *Water Power* 31(9) (2005)
5. Grey, C.y.: System Theory and Cloud Model Based Anti-accuracy Analysis of Flood Disater. *Huazhong University of Science and Technology, Hubei* (June 2010)

Research on Cut Order Planning for Apparel Mass Customization

Liu Yan-mei^{1,2}, Yan Shao-cong², and Zhang Shu-ting²

¹ Institute of Contemporary Manufacturing Engineering, Zhejiang University,
Hangzhou 310027, China

² Faculty of Mechanical Engineering & Automation, Zhejiang Sci-Tech University,
Hangzhou 310018, China

Abstract. To solve the problem of versatile sizes and irregular number in cut order planning (COP) for apparel mass customization, the mathematical model was built and optimization method based on probability search was proposed. Several cutting table layout plans were generated randomly with the production constriction. The optimized sizes combination plan was accordingly obtained using probability search algorithm. The optimization method could rapidly get the apparel cutting plan and decrease the number of cutting table.

Keywords: Mass customization, apparel, cut order planning, optimization, probability search.

1 Introduction

In current market circumstances, orders apparel enterprises received are generally versatile, middle or small batch [1]. Apparel enterprises are faced with a major challenge of quick response to customers' needs such as figure type, style and color priority, sizes, number and so on [2]. Apparel cut order planning(COP) mainly determines the number of cutting table, the sizes combination and the lay according to production capacity. COP can not only give the cutting solution but also make use of the production capacity, save fabric and labor and increase profit.

In current studies, part of the research methods in essence was based on artificial experience which was more convenient to solve problems with less sizes and numbers [3]. However the algorithms were difficult to expand to mass customization apparel production with more sizes and numbers. In addition, part of the above studies was constructive and difficult to deal with the above situation[4]. Furthermore some algorithms yielded bigger searching space leading to low efficiency and long calculating time [5].

In response to these problems, this paper proposes optimization method based on probability search. Several initial cutting table layout plans were generated randomly with the production constriction. The optimized sizes combination plan was accordingly obtained using probability search algorithm. The optimization method could rapidly get the apparel cutting plan and decrease the number of cutting table.

2 Optimization Model of COP

Apparel orders varies according to market or customer needs including style, color, size, quantity, material, brand and other attributes and at least color, size, number these three attributes. Take the simplest case for example, an apparel company received the order of single style, single fabric, the same color and multiple sizes, such as official's uniform. The order was shown in Table 1.

Table 1. Apparel Order for mass customization

size	<i>l</i>	<i>2</i>	...	<i>i</i>	...	<i>m</i>
number	Y_1	Y_2	...	Y_i	...	Y_m

In certain production circumstance, the largest layer of the fabric is constant P_{max} and the maximum pieces of clothes each cutting table can cut is constant c . The constant P_{max} is mainly decided by the length of cut blade and the thickness of fabric. The constant c is decided by the size of the clothes and cutting table. It is assumed that n cutting tables should be needed to complete the order. Denote the i size are cut in the j cutting tables, the pieces of clothes is a_{ij} , layers of cloth cut on the j cutting table is X_j , where $0 \leq a_{ij} \leq c$.

The objective are $\min(n)$. It means in actual production the total number of pieces and cutting table should be minimum to meet customer's demand.

Constrictions are

$$0 \leq \sum_{i=1}^m a_{ij} \leq c. \tag{1}$$

$$X_j \leq P_{max}. \tag{2}$$

$$Y_i \leq \sum_{j=1}^n a_{ij} X_j, j = 1, 2, \dots, n. \tag{3}$$

Formula (1) means that cut the number of clothes cut on each cutting table could not exceed the maximum number. Formula (2) expresses that the number of spreading layer could not exceed the maximum number. Formula (3) shows that the number of clothes cut for each size must exceed the number of orders for the same size. The formulas would be transformed into matrix as shown in formula (4).

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_i \\ \vdots \\ Y_m \end{bmatrix} \leq \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mj} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_j \\ \vdots \\ x_n \end{bmatrix} \quad (4)$$

In formula (4) matrix X means spreading layers on each cutting table. Matrix a means the sizes on each cutting table. So a COP solution could be obtained with the matrix X and a . Since both matrix X and a are unknown, what's more, matrix X and a are matched, inversion matrix method could not be used to get results. The problem is essentially an integer programming problem, but also a NP problem. There is no tool with excellent performance to support the solution. Therefore, the paper proposed two stage optimization method based on probability search and genetic algorithm to solve it.

3 Optimization Method Based on Probability Search for COP

The optimization method based on probability search means a number of initial spreading solutions are randomly generated within certain constraints. Accordingly the sizes combination solution could be obtained using searching algorithm combined with probability.

3.1 Searching for the Spreading Solution on Cutting Tables

Spreading solution on cutting tables means to determine the number of cutting tables and the spreading layers on each table. For an apparel order these two values interact. When the number of cutting table changes, the spreading layers on each table will change accordingly. These two variables have multiple sets of feasible solutions, and in the initial algorithm stage it is difficult to get the exact value. Therefore, random searching method was used to solve the problem.

1、The number of cutting tables n

The number of cutting tables could be searched in the natural number range. Considering reducing the searching space and improving the efficiency of the algorithm in the actual calculation process, the number of cutting tables was defined a range of values. The range was jointly determined by the number of apparel order and the maximum cutting capacity of the cutting table.

2、The spreading layers on each cutting table X_j

Randomly generate spreading layers X_j on each cutting table of totally n , and in accordance with formula (2).

3.2 Sizes Combination Solution Optimization Based on Probability Search

Size combination solution could be obtained by search algorithm with probability. Based on the maximum number of clothes and the spreading solution, the size combination solution could be randomly generated to meet the apparel order. The detailed calculation method was the following.

$$Y_i - \sum_{j=1}^{k-1} a_{ij} x_j$$

Step 1. Calculate $\max a_{ij} = \frac{Y_i - \sum_{j=1}^{k-1} a_{ij} x_j}{x_j}$, if $\max a_{ij} \leq c$, the clothes of size

i arranged on cutting table j was a_{ij} , which was randomly generated between the integer of 0 and $\max a_{ij}$ with the same probability. If $\max a_{ij} > c$, a_{ij} was randomly generated between the integer of 0 and c .

Step 2. Calculate $\sum_{j=1}^n a_{ij} X_j$, if $\sum_{j=1}^n a_{ij} X_j < Y_i$, reduce i units from the current cutting table number, then re-run step 1, generate a new a_{ij} , and the initial value of i was 1.

Step 3. If the regenerated a_{ij} up to i cutting tables would still be unable to meet the constraints of formula (1), add 1 to i , re-run step 3, and generate new a_{ij} .

Step 4. Until $i = m$, get matrix a .

Several cut table spreading solutions and sizes combination solutions were obtained. Additional clothes would be produced according to the spreading and sizes matching solution obtained in the first stage. These additional clothes would increase the inventory, fabric and labor cost. It could not meet the actual production demand. Therefore, it is necessary to optimize the solution. The number of the clothes of over production could be controlled to the extent permitted in the enterprise.

4 Conclusion

The paper researched the COP problem with versatile sizes and irregular number of clothes for apparel mass customization. The mathematical model of the problem was built and optimization method based on probability search was proposed. Subsequent research could focus on how to solve the COP problem of more customized apparel order, as well as overall optimization of production plans combined with other aspects of clothing production[6]. In addition, other optimization methods could be used for further research in apparel production plan[7].

Acknowledgments. This work is supported by the National Key Technology R&D Program, China(No. 2011BAB02B01) and Science Foundation of Zhejiang Sci-Tech University (ZSTU) under Grant No.111329A4Y10810.

References

1. Jacobs-Blecha, C., Ammons, J.C., Schutte, A., Smith, T.: Cut order planning for apparel manufacturing. *IIE Transactions* 30, 79–90 (1998)
2. Dawn, M.R., Douglas, R.S.: Cut scheduling in the apparel industry. *J. Computers & Operations Research* 34, 3209–3228 (2006)
3. Jiang, X.-w.: Research on apparel cut order planning optimization system. *Sichuan Silk* 12(2), 41–44 (2000)
4. Sun, X.-y., Qiu, Y.-y.: How to make cut order planning in apparel production. *China Textile Leader* 2, 40–41 (2003)
5. Bouziri, A., M'Hallah, R.: A Hybrid Genetic Algorithm for the Cut Order Planning Problem. In: Okuno, H.G., Ali, M. (eds.) *IEA/AIE 2007. LNCS (LNAI)*, vol. 4570, pp. 454–463. Springer, Heidelberg (2007)
6. Wong, W.K., Leung, S.Y.S.: Genetic optimization of fabric utilization in apparel manufacturing. *Int. J. Production Economics* 114, 376–387 (2008)
7. Degraeve, Z., Vandebroek, M.: A mixed integer programming model for solving a layout problem in the fashion industry. *Management Science* 44, 301–310 (1998)

The Semantics of Dynamic Fuzzy Logic Programming Language

Xiaofang Zhao

School of Computer Science & Technology, Shandong Institute of Business and Technology
Yantai, China
xf_zh@163.com

Abstract. Dynamic fuzzy problems exist extensively in realistic world. The dynamic fuzzy logic (DFL) programming language is to deal with dynamic fuzzy data. In order to implement DFL programming language, it should be firstly defined. In this paper, we give the denotational semantics of DFL programming language. The work mainly includes modifying the classical lambda calculus to introduce the character of dynamic fuzzy, the descriptions of semantic objects and the handling functions of semantics.

Keywords: Dynamic fuzzy logic (DFL), Programming language, Denotational semantics.

1 Introduction

Events having the character of dynamic fuzzy logic exist extensively. Such as stock market index, economic growth, weather changes and so on. In order to simulate coping with dynamic fuzzy problems in computer, it is necessary to provide software, that is, to research and design a programming language. There have been researches on fuzzy programming language to deal with fuzzy data, such as the references [1,2,3]. And there have been researches on dynamic programming language to deal with dynamic data, such as the reference [4]. Up to now, there has been few programming language to deal with dynamic fuzzy data. In reference [5], we have attempted to give the frame of DFL programming language. In this paper, we define the DFL programming language from the point of denotational semantics. Denotational semantics is based on more mature mathematics theory such as lambda calculus, domains theory and so on. To deal with dynamic fuzzy data, the traditional lambda calculus should be changed.

This paper is organized as follows: In the section 2, the conception of dynamic fuzzy data has been introduced. In section 3, some basic knowledge about lambda calculus has been given. In section 4, the denotational semantics of DFL programming language has been described in detail. In section 5, the conclusion has been presented.

2 Dynamic Fuzzy Data

Pact 1. The character of data with both dynamic and fuzzy is called dynamic fuzzy character.

Eg.1 The global climate becomes more and more warming.

The word "becomes" reflects the dynamic character and "warming" reflects the fuzzy character.

Pact 2. The data with the dynamic fuzzy character is called dynamic fuzzy data.

We call the whole clause mentioned by Eg.1as dynamic fuzzy data.

The sets of dynamic fuzzy datum i.e. dynamic fuzzy sets are defined as follows:

Definition 1. Let a mapping be defined in the domain U.

$$(\overleftarrow{A}, \overrightarrow{A}) : (\overleftarrow{U}, \overrightarrow{U}) \rightarrow [0,1]^\times [\leftarrow, \rightarrow], (\overleftarrow{u}, \overrightarrow{u}) \mapsto (\overleftarrow{A}(\overleftarrow{u}), \overrightarrow{A}(\overrightarrow{u}))$$

We write $(\overleftarrow{A}, \overrightarrow{A}) = \overleftarrow{A}$ or \overrightarrow{A} , then we name $(\overleftarrow{A}, \overrightarrow{A})$ the dynamic fuzzy sets (DFS) of $(\overleftarrow{U}, \overrightarrow{U})$ and name $(\overleftarrow{A}(\overleftarrow{u}), \overrightarrow{A}(\overrightarrow{u}))$ the membership degree of membership function to $(\overleftarrow{A}, \overrightarrow{A})$.

Dynamic fuzzy variable is usually symbolized by $(\overleftarrow{x_1}, \overrightarrow{x_1}), (\overleftarrow{x_2}, \overrightarrow{x_2}), \dots$ or $(\overleftarrow{x}, \overrightarrow{x}), (\overleftarrow{y}, \overrightarrow{y}), \dots$ where $[\leftarrow, \rightarrow]$ indicates the direction of dynamic change.

3 Lambda Calculus

Lambda calculus is based on function calculus. It is suitable for the description of semantics especially for denotational semantics. With the aim of describing the denotational semantics of dynamic fuzzy logic programming language, we need to modify the classical lambda calculus to introduce the character of dynamic fuzzy. From the definition in the section 2, we can see that to introduce the dynamic fuzzy character we need to add a special set $D=[0,1]^\times [\leftarrow, \rightarrow]$, where " \leftarrow "denotes good or advance direction of dynamic change and " \rightarrow "indicates bad or back direction of dynamic change.

To introduce the set D into the classical lambda calculus, a simple method is that terms are represented by trees whose branches labeled with elements of D. In detail, our terms are composed of two parts: one part belongs to the classical lambda calculus and the other part is a list built on the set D. Therefore, for each term M we can write it as $[m, l(M)]$. Thus term should be redefined.

Definition 2. Terms can be recursively defined as the follows:

If $(\overleftarrow{x}, \overrightarrow{x})$ is a variable then for each $\alpha \in D$, $[\overleftarrow{x}, \overrightarrow{x}], \alpha$ is a term.

If $M=[m, l(M)]$ is a term and $(\overleftarrow{x}, \overrightarrow{x})$ is a variable then for each $\alpha \in D$, $\lambda (\overleftarrow{x}, \overrightarrow{x}).M = [\lambda (\overleftarrow{x}, \overrightarrow{x}).m, \alpha \cdot l(M)]$ is a term.

If both $M=[m, l(M)]$ and $N=[n, l(N)]$ are terms, then for each $\alpha \in D$, $MN = [m \ n, \alpha \cdot l(M) \cdot l(N)]$ is a term.

Compared with terms of the classical lambda calculus we can see that in our transformed lambda calculus a variable isn't a term any more unless it is labeled by an element of set D.

Below we show an example about terms, see as example Eg.2.

Eg.2 Let $m = \lambda \overset{\leftrightarrow}{xy} . (\bar{x}\bar{y} + \bar{x}) (\lambda \bar{z} . \bar{z})$, where m belongs to the classical lambda calculus part of the term M, and $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8, \alpha_9, \alpha_{10} \in D$ which are to label m, then M can be write as:

$$[\lambda \overset{\leftrightarrow}{xy} . (\bar{x}\bar{y} + \bar{x}) (\lambda \bar{z} . \bar{z}), \alpha_1 \cdot \alpha_2 \cdot \alpha_3 \cdot \alpha_4 \cdot \alpha_5 \cdot \alpha_6 \cdot \alpha_7 \cdot \alpha_8 \cdot \alpha_9 \cdot \alpha_{10}]$$

and its structure as a tree is the following figure Fig.1:

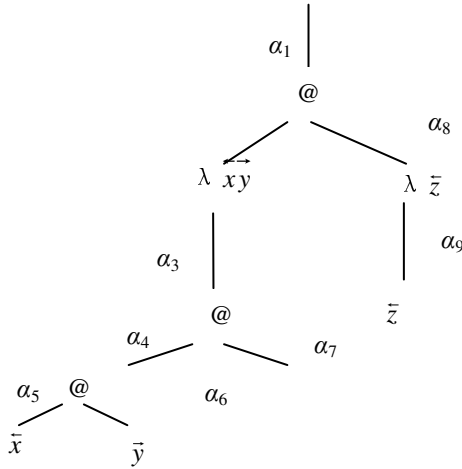


Fig. 1. The structure of Eg.2

Let's analyze above structure: The dynamic fuzzy membership degree of \bar{x} is α_5 and the dynamic fuzzy membership degree of \bar{y} is α_6 . While the dynamic fuzzy membership degree of the product of \bar{x} and \bar{y} is $\alpha_4 = t(\alpha_5, \alpha_6)$ which is different from both α_5 and α_6 where t is T modular operation [6] or S modular operation [6] of DFS. The rest can be gotten by analogy.

4 The Denotational Semantics of DFL Programming Language

The denotational semantics means making each program element correspond to a mathematical object that called a denotation of its corresponding program element and having each program a mapping from input field to output field where the input field and output field both called domain of discourse. Then we can see that the description of denotational semantics must be based on some mathematics theory such as lambda calculus, domain theory and so on. In general the description of denotational semantics is composed of four parts, namely, the abstract syntax, the

conditions of context, the descriptions of semantic objects and the handling functions of semantics. The abstract syntax infers giving the syntactic categories. The conditions of context can example mistakes of syntax. The semantic objects are the mathematical objects mapped from program elements. When describing semantic objects, it means to give the semantic categories. The handling function of semantics can simply be described as mappings from syntactic categories to their corresponding semantic categories. Among the four parts the handling function of semantics is the most important. Below we give the denotational semantics of DFL programming language in detail.

4.1 Abstract Syntax

In this part we give the syntactic structure of the DFL programming language following the guarded commands, see [7], proposed by Dijkstra on account of its characteristic that is to introduce nondeterminism to a program each sentence possibly executed is provided a guarded condition.

$$\begin{aligned}
 &S ::= \text{skipl} \\
 &\quad \text{abortl} \\
 &\quad (\vec{v}, \vec{v}) = \text{El} \\
 &\quad (S; S) \mid \\
 &\quad \text{if } G \text{ fl} \\
 &\quad \text{do } G \text{ od} \\
 &E ::= (\vec{v}, \vec{v}) \mid \\
 &\quad (\vec{n}, \vec{n}) (\vec{d}, \vec{d}) \mid \\
 &\quad E \text{ op } E \\
 &G ::= B \rightarrow S \mid \\
 &\quad (G \square G) \\
 &B ::= (\text{true}, (\vec{d}, \vec{d})) \mid \\
 &\quad (\text{false}, (\vec{d}, \vec{d})) \mid \\
 &\quad B \text{ Bop } B \mid \\
 &\quad E \text{ rel } E
 \end{aligned}$$

With the following syntactic categories:

$S \in$ sentences

$E \in$ expressions

$G \in$ guarded commands

$B \in$ boolean expressions

$(\vec{v}, \vec{v}) \in$ variables

$(\vec{n}, \vec{n}) \in$ constants

$(\vec{d}, \vec{d}) \in D$, indicates membership degree

op indicates operation

Bop indicates boolean operation

rel indicates relation operation

4.2 The Conditions of Context

In our case the conditions of context are merged into the handling functions of semantics. When finding mistakes the value is assigned to “error”. It doesn’t detailedly explain how to deal with mistakes.

4.3 The Descriptions of Semantic Objects

According to the abstract syntax given above, we give the basic syntactic categories of DFL programming language as follows:

Literals :

Positive Numbers Domain Num+ = {0,1,2,..... }
 Negative Numbers Domain Num- = {-1,-2,-3,..... }
 Integers Domain Int = Num ⊕ Num-
 Booleans Domain Bool = {True, False}
 Dynamic Fuzzy Greeds Domain Grd = [0,1] × [←,→]
 Values Domain Val = Int ⊕ Bool ⊕ Grd

Dynamic Fuzzy Variables Domain DFLVar

Dynamic Fuzzy Constants Domain DFLCon

Dynamic Fuzzy Integers Domain DFLInt

DFLInt=Int × Grd

Dynamic Fuzzy Booleans Domain DFLBool

DFLBool=Bool × Grd

Dynamic Fuzzy Values Domain DFLValue

DFLValue=DFLInt ⊕ DFLBool

Dynamic Fuzzy Identifiers Domain DFLIde

Dynamic Fuzzy Identifier Values Domain DFLIdeVal

DFLEdeVal=DFLInt

Dynamic Fuzzy Store Domain DFLSto

DFLSto=DFLVar → DFLValue

Dynamic Fuzzy Environment Domain DFLEnv

DFLEnv = DFLVar → DFLIdeVal

4.4 The Handling Function of Semantics

The basic formation of the handling function of semantics is as follows:

$$E: \text{Exp} \rightarrow U \rightarrow S \rightarrow E$$

Where Exp is syntactic category and $U \rightarrow S \rightarrow E$ is semantic category. The detailed formation is the following: The syntax is bracketed by the symbol of “[]” and the semantic parameter is written behind of “[]”. The symbol of “=” is between of them. Below we give the handling functions of semantics of DFL programming language.

$S : S \rightarrow DFLEnv \rightarrow DFLSto \rightarrow S (DFLSto)$

$S [[\text{skip}]] = \lambda(\sigma, (\bar{d}, \bar{d})). (\{\sigma\}, (\bar{d}, \bar{d}))$

$S [[\text{abort}]] = \lambda(\sigma, (\bar{d}, \bar{d})). (\{\delta\}, (\bar{d}, \bar{d}))$

$$S [[(\vec{v}, \vec{v}) = E]] = \lambda(\sigma, (\vec{d}, \vec{d})). ((DFLSto[[(\vec{v}, \vec{v})]] \\ (E[[E]] \sigma)(\sigma, (\vec{d}, \vec{d}))), (\vec{d}, \vec{d}))$$

$$S [[(S1; S2)]] = \text{ext}(S[[S2]] S[[S1]])$$

$$S [[\text{if } G \text{ fi}]] = \lambda(\sigma, (\vec{d}, \vec{d})).$$

$$((\vec{d}, \vec{d}), \text{on2}[\lambda(\vec{x}, \vec{x}) \in O. \{|\delta|\}] G [[G]](\sigma, (\vec{d}, \vec{d})))$$

$$S [[\text{do } G \text{ od}]] = \text{fix}([\lambda f \lambda(\sigma, (\vec{d}, \vec{d})). ((\vec{d}, \vec{d}), \\ \text{on2}[\lambda(\vec{x}, \vec{x}) \in O. \{|\sigma|\}, \text{ext}(f)] \\ (G [[G]](\sigma, (\vec{d}, \vec{d}))))])$$

$E: \text{Exp} \rightarrow DFLEnv \rightarrow DFLSto \rightarrow DFLInt$

$$E [[(\vec{v}, \vec{v})]] = \lambda\sigma. DFLSto[[(\vec{v}, \vec{v})]]\sigma$$

$$E [[(\vec{n}, \vec{n}) (\vec{d}, \vec{d})]] = \lambda\sigma. [(\vec{n}, \vec{n}) (\vec{d}, \vec{d})]$$

$$E [[E1 \text{ op } E2]] = \lambda\sigma. \text{let} [(\vec{n}_1, \vec{n}_1) (\vec{d}_1, \vec{d}_1)] \\ = E [[E1]]\sigma \text{ in } \text{let} [(\vec{n}_2, \vec{n}_2) (\vec{d}_2, \vec{d}_2)] \\ = E [[E2]]\sigma \text{ in } [(\vec{n}_1, \vec{n}_1) \text{ op } (\vec{n}_2, \vec{n}_2), s((\vec{d}_1, \vec{d}_1), (\vec{d}_2, \vec{d}_2))]$$

$G: G \rightarrow DFLEnv \rightarrow DFLSto \rightarrow DFLBool$

$$G [[B \rightarrow S]] = \lambda\sigma. B [[B]]\sigma$$

$$G [[G1 \square G2]] = \lambda\sigma. (G [[G1]]\sigma \wedge G [[G2]]\sigma)$$

$B: BExpr \rightarrow DFLEnv \rightarrow DFLSto \rightarrow DFLBool$

$$B [[(\vec{true}, \vec{true})(\vec{d}, \vec{d})]] = \lambda\sigma. [(\vec{true}, \vec{true})(\vec{d}, \vec{d})]$$

$$B [[(\vec{false}, \vec{false})(\vec{d}, \vec{d})]] = \lambda\sigma. [(\vec{false}, \vec{false})(\vec{d}, \vec{d})]$$

$$B [[B1 \text{ Bop } B2]] = \lambda\sigma. \text{let} [b1, (\vec{d}_1, \vec{d}_1)] = B [[B1]]\sigma \text{ in } \text{let} [b2, (\vec{d}_2, \vec{d}_2)] \\ = B [[B2]]\sigma \text{ in } [(\vec{n}_1, \vec{n}_1) \text{ Bop } (\vec{n}_2, \vec{n}_2), s((\vec{d}_1, \vec{d}_1), (\vec{d}_2, \vec{d}_2))]$$

Where s is T modular operation[6] or S modular operation[6] of DFS.

5 Example

Below we show a simple program to observe the process of dealing with dynamic fuzzy data.

DFLexample

$$\{\text{DFInt } \vec{x}_1 = (\vec{6}, \vec{0.7}), \vec{x}_2 = (\vec{12}, \vec{0.8});$$

$$\text{DFInt } \vec{x}_3 = (\vec{4}, \vec{0.6}), \vec{x}_4 = (\vec{2}, \vec{0.5});$$

$$\text{DFReal } \vec{x}_5 = (\vec{2.5}, \vec{0.7}), \vec{x}_6;$$

DO

\ \ Sqrt() denotes computing square root

$$\vec{x}_6 = \text{SQRT}(\vec{x}_2 * \vec{x}_2 - 4 * \vec{x}_1 * \vec{x}_3)$$

```

IF  $\vec{x}_6 > \overleftarrow{x}_5$ 
    Then  $\vec{x}_2 = \vec{x}_2 \cdot \overleftarrow{x}_4$ 
FI
OD
}
    
```

Results:

From the denotational semantics above given we see the value of dynamic fuzzy membership degree will be changed through T modular operation or S modular operation during the execution of the program. There are several forms of T modular and S modular operation, but here we only take example for two kinds of them.

Table 1. The result of DFLExample

Name of	stateσ1	stateσ2	stateσ3
t(a , b)=min(a , b)			
\vec{x}_1	$(\vec{6}, \vec{0.7})$	$(\vec{6}, \vec{0.7})$	$(\vec{6}, \vec{0.7})$
\vec{x}_2	$(\vec{12}, \vec{0.8})$	$(\vec{10}, \vec{0.5})$	$(\vec{10}, \vec{0.5})$
\vec{x}_3	$(\vec{4}, \vec{0.6})$	$(\vec{4}, \vec{0.6})$	$(\vec{4}, \vec{0.6})$
\vec{x}_4	$(\vec{2}, \vec{0.5})$	$(\vec{2}, \vec{0.5})$	$(\vec{2}, \vec{0.5})$
\overleftarrow{x}_5	$(\vec{2.5}, \overleftarrow{0.7})$	$(\vec{2.5}, \overleftarrow{0.6})$	$(\vec{2.5}, \overleftarrow{0.6})$
\vec{x}_6		$(\vec{6.9}, \vec{0.6})$	$(\vec{2}, \vec{0.5})$
t(a , b)=max (0,a+ b-1)			
\vec{x}_1	$(\vec{6}, \vec{0.7})$	$(\vec{6}, \vec{0.7})$	$(\vec{6}, \vec{0.7})$
\vec{x}_2	$(\vec{12}, \vec{0.8})$	$(\vec{10}, \vec{0.3})$	$(\vec{10}, \vec{0.3})$
\vec{x}_3	$(\vec{4}, \vec{0.6})$	$(\vec{4}, \vec{0.6})$	$(\vec{4}, \vec{0.6})$
\vec{x}_4	$(\vec{2}, \vec{0.5})$	$(\vec{2}, \vec{0.5})$	$(\vec{2}, \vec{0.5})$
\overleftarrow{x}_5	$(\vec{2.5}, \overleftarrow{0.7})$	$(\vec{2.5}, \overleftarrow{0.7})$	$(\vec{2.5}, \overleftarrow{0.7})$
\vec{x}_6		$(\vec{6.9}, \vec{0})$	$(\vec{2}, \vec{0.})$

6 Conclusion

Compared to relevant work, what is the characteristic in this paper is that we combined the character of dynamic with the character of fuzzy for research but not only the character of fuzzy as the reference [3] or only the character dynamic as

reference [4] and described the semantics of DFL programming language in the terms of denotational semantics. What we have done in this paper is summarized as follows:

We have modified the classical lambda calculus to introduce a special set D whose elements will be used to label its terms.

We have given denotational semantics of DFL programming language which include abstract syntax, the descriptions of semantic objects and the handling functions of semantics.

References

1. Adamo, J.M.: L.P.L. A Fuzzy Programming Language: 1. syntactic aspects. *J. Fuzzy Set and Systems* 3, 151–179 (1980)
2. Adamo, J.M.: L.P.L. A Fuzzy Programming Language: 2. Semantic Aspects. *J. Fuzzy Set and Systems* 3, 261–289 (1980)
3. Alvarez, D.S., Gó mez Skarmeta, A.F.: A fuzzy language. *J. Fuzzy Sets and System* 141, 335–390 (2004)
4. Tang, Z.-s.: *Temporal Logic programming and software engineering* (in Chinese). Science Press, Beijing (2002)
5. Zhao, X.: The Frame of DFL Programming Language. In: *7th International Conference on Fuzzy Systems And Knowledge Discovery*, Hai Nan, China, pp. 343–348 (2010)
6. Li, F.-z., Zheng, J.-l.: Module Operation of Dynamic Fuzzy Stes. *Journal of the Central University for Nationalities* 16, 96–101 (1997) (in Chinese)
7. Lu, R.-q.: *Formal semantics of computer languages*. Science Press, Beijing (1992) (in Chinese)

Trend Prediction of Oil Temperature for Wind Turbine Gearbox Based on Grey Theory

Wang Rui and Li Gang*

Taiyuan University of Technology
Taiyuan, China
{wangrui01, ligang}@tyut.edu.cn

Abstract. Considering the stochastic volatility of the wind turbine gearbox oil temperature, the wavelet packet is used to eliminate its noise. On this basis, the grey model is applied to forecast the wind turbine gearbox oil temperature. The predicted results show that the wavelet packet and the grey prediction method have better forecast accuracy. The wind turbine gearbox oil temperature trends can be predicted timely and accurately.

Keywords: Wind Turbine, Gearbox Oil Temperature, Grey Model, Wavelet Packet De-noising, Trend Prediction.

1 Introduction

Wind turbine is a complex nonlinear uncertain system, its safe and reliable operation has a direct impact on the stability of interconnected power grid. In order to ensure the security of grid operation and improve system quality, it is important to detect early the failures of wind turbine [1~4].

In the operation of wind turbine gearbox, due to the surface friction, heat radiation and other reasons, the temperature variation of the gearbox is more complicated. It is difficult to find its mathematical model. For obtaining the conditions and trends of the wind turbine gearboxes operating temperature timely, based on the actual monitoring data of wind turbine gearbox oil temperature in a wind farm, the wavelet packet is used to de-noise its noise, and the grey theory is adopted to predict its trend.

2 Wavelet Packet De-noising Method of Oil Temperature for Wind Turbine Gearbox

For the sake of analyze better the trends of the wind turbine gearbox oil temperature, it is essential to remove the random fluctuations of the oil temperature time series. In this paper, the wavelet packet method [5] is used to eliminate the noise of wind turbine gearbox oil temperature.

There are two key points of the wavelet packet de-noising methods: the selection of the threshold and the threshold value quantizing of the wavelet packet coefficients.

* Corresponding author.

They are directly related to signal de-noising quality to a certain extent. The steps of the wavelet packet transform de-noising as follow: the wavelet packet decomposition of the signal, the determination of the optimum wavelet packet base, the threshold value quantizing of the wavelet packet coefficients, the reconstruction of wavelet packet signal. The detailed steps can be seen in [5].

3 Grey Prediction Theory

The grey system theory [6] said that all random processes are changes in a certain amplitude range and a certain amount of time zone. The random process is called grey process. In fact many historical data has shown a significant index rule after a cumulative process. And most systems are generalized energy system, the index change law is a law of energy. Grey prediction is used to predict the trends of wind turbine gearbox oil temperature. The AGO got a strong regular sequence, and then exponential curve fitting and forecasting, and finally by reduction with forecasts by tired value. Grey forecasting model is a basic model fitting parameters, namely, GM (1, 1). It is through the accumulation of raw data generated, got a strong regular sequence, and then fitting an exponential curve to the predicted value.

The GM (1, 1) model can be defined as:

$$x^{(0)}(k) + az^{(1)}(k) = b \tag{1}$$

where

$$z^{(1)}(k) = 0.5[x^{(1)}(k) + x^{(1)}(k - 1)] \tag{2}$$

For the original data sequence

$$X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\} \tag{3}$$

where n is the number of data.

The steps of GM(1,1) model as follows:

Step 1: The first-order accumulated generating operation (1-AGO) of X(0) is given as

$$x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i), \quad k = 1, 2, \dots, n. \tag{4}$$

$$X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)\} \tag{5}$$

Step2: Mean generation with consecutive neighbors of X(1).

The generated mean sequence Z(1) of X(1) is defined as:

$$Z^{(1)} = \{z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n)\} \tag{6}$$

where $z^{(1)}(k) = 0.5[x^{(1)}(k) + x^{(1)}(k - 1)]$, $k=2, 3, \dots, n$.

and

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix} \tag{7}$$

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix} \tag{8}$$

Step3: Least squares method estimate the parameter series $\hat{a} = \begin{bmatrix} a \\ b \end{bmatrix}$.

$$\hat{a} = (B^T B)^{-1} B^T Y \tag{9}$$

Step4: The GM(1,1) model can be constructed by establishing a first order differential equation for $x^{(1)}(k)$ as:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \tag{10}$$

and response formula:

$$\hat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{b}{a} \right) e^{-ak} + \frac{b}{a} \tag{11}$$

Step5: Generating predicted value

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \tag{12}$$

Step6: Applying the residual formula

$$e^{(0)}(k) = \frac{|x^{(0)}(k) - \hat{x}^{(0)}(k)|}{x^{(0)}(k)} \times 100\% \tag{13}$$

to test the model accuracy. If the relative error $e^{(0)}(k)$ is smaller, then indicates the model has higher prediction accuracy. If the test is unqualified, then amend the original model through establishing residual GM (1,1) model.

4 Oil Temperature Trend Prediction of Wind Turbine Gearbox Based on Wavelet Packet Filtering and Grey Model

To ensure the safe and stable operation of wind turbine gear box, estimate the performance and trend of gearbox timely, the data of wind turbine gear box oil temperature from a wind farm are selected. The singularity points are removed using wavelet packet. Ten minutes sampling points for the data, 450 points are selected to analyze. Fig. 1 is the data of wind turbine gearbox oil temperature, it can be seen from Fig. 1, the wind turbine itself is complex, but also by the multi-load cross-impact, making the gearbox oil temperature time series is very complicated, there is more strong volatility, difficult to accurately analyze the figure its regularity. For removing the random fluctuations and improving predicted veracity, it is need to de-noise its noise. Fig. 2 is the result of removing the singular point, reducing noise of the wind turbine gearbox oil temperature time series data. It can be seen from Fig. 2, after wavelet packet filtering, the trend curve of wind turbine gearbox oil eliminates the volatility of the gearbox oil temperature, and it can reflect the trend of gearbox oil temperature.

In this paper, the 100 ~ 200 point data of gearbox oil temperature are used to predict its trend. Firstly, applying 100 ~ 107-point data to predict the 108-point using equations (4) ~ (12). Secondly, supplement the latest data corresponding to 108-point, at the same time removing the oldest data (100-point). Then re-establish GM (1,1) model to predict the next value, and the latest corresponding data added, while removing the oldest data, and so on. And calculate the relative error $e(0)(k)$ using residual formula (13).

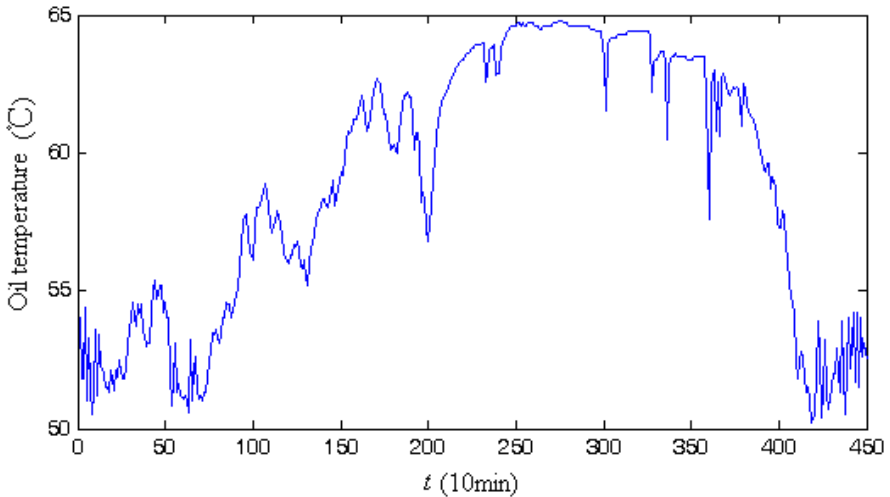


Fig. 1. The original data of gearbox oil temperature

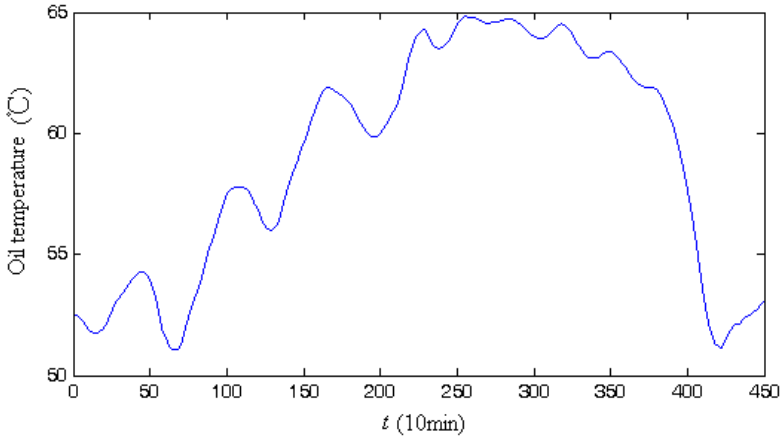


Fig. 2. The de-noised data of gearbox oil temperature using spatial correlation filtering method

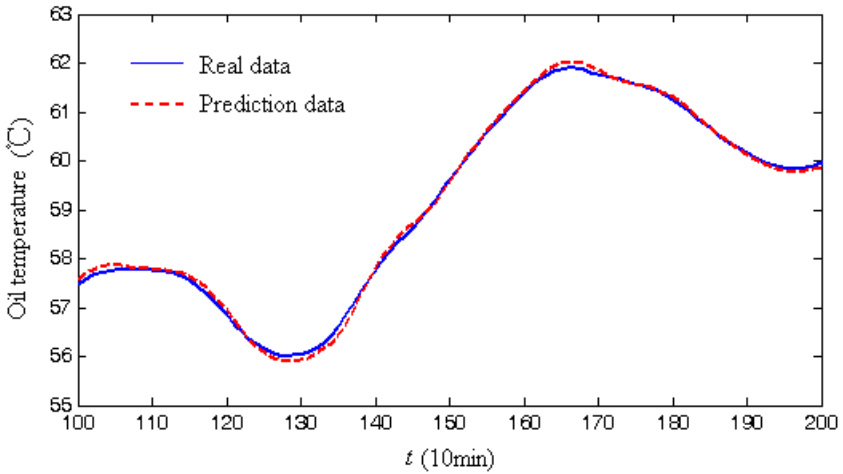


Fig. 3. The trend prediction results of gearbox oil temperature

Figs. 3 and 4 is respectively the trend prediction curve and the forecast error curve of gearbox oil temperature based on grey model. From Fig. 3, it can be seen, the application of grey model to predict trend of oil temperature for wind turbine gearbox, has higher accuracy, the predictive value and the actual values are basically consistent, the average relative error is 0.12%. At 136x10min moment, the relative prediction error is the largest with 0.31%.

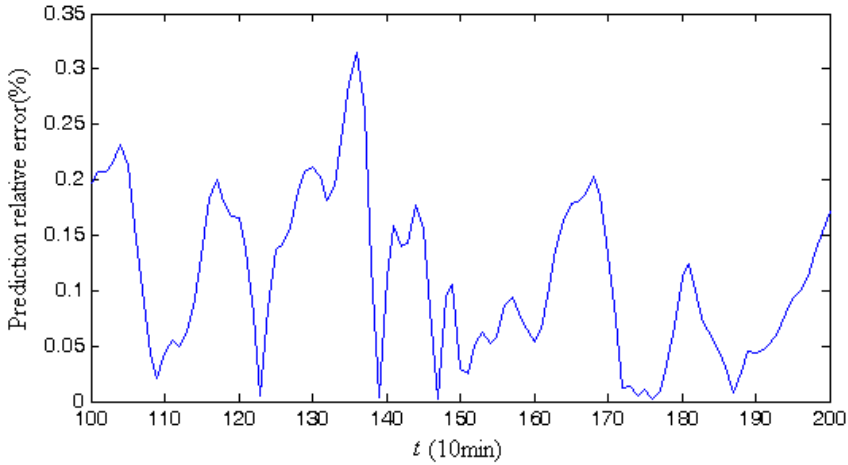


Fig. 4. The prediction relative error of gearbox oil temperature

5 Conclusion

In the paper, wavelet packet filtering method is used to de-noise the noise of oil temperature for wind turbine gearbox, the random fluctuations of oil temperature time series is eliminated. The grey model is applied to predict gearbox oil temperature trend after pretreating the wind turbine time series data. Prediction results show that this method has good prediction accuracy, can timely and accurately predict the development trend of oil temperature for wind turbine gearbox, and has some practical value.

References

- [1] An, X., Jiang, D.: Chaotic characteristics identification and trend prediction of running state for wind turbine. *Electric Power Automation Equipment* 30(3), 15–19 (2010) (in Chinese)
- [2] Dong, L., Wang, L., Gao, S., et al.: Modeling and Analysis of Prediction of Wind Power Generation in the Large Wind Farm Based on Chaotic Time Series. *Transactions of China Electrotechnical Society* 23(12), 125–129 (2008) (in Chinese)
- [3] Luo, H., Liu, T., Li, X.: Chaotic Forecasting Method of Short-Term Wind Speed in Wind Farm. *Power System Technology* 33(9), 67–71 (2009) (in Chinese)
- [4] Zhang, G., Zhang, B.: Wind Speed and Wind Turbine Output Forecast Based on Combination Method. *Automation of Electric Power Systems* 33(18), 92–95 (2009) (in Chinese)
- [5] Bettayeb, F., Haciane, S., Aoudia, S.: Improving the time resolution and signal noise ratio of ultrasonic testing of welds by the wavelet packet. *NDT & E International* 38, 478–484 (2005)
- [6] Deng, J.L.: *Grey Prediction & Decision*. Huazhong Engineering College Press, Wuhan (1985)

The Semantics of Logic System $lp(X)$

Hua Li

Department of Information Engineering, Hanzhou Polytechnic College, Hanzhou 311402,
Zhejiang, P.R. China
zjlihua@126.com

Abstract. Lattice-valued logic plays an important role in Multi-valued logic systems. In this paper, a lattice valued logic system $lp(X)$ is constructed. The semantics of $lp(X)$ is discussed. It may be more convenient in application and study especially in the case that the valuation domain is finite lattice implication algebra.

Keywords: Semantic, Logic, Lattice, Logic System, Lattice Implication Algebra.

1 Introduction

Multi-valued logic is an extension and a development of classical logic. Lattice-valued logic is an important case of multi-valued logic because a lattice is a kind of important algebraic structure that is an abstract model of the real world. Lattice-valued logic extends the chain-type truth-valued field to a general lattice, which can more efficiently model the uncertainty of human thinking, judgment and decision-making.

In many famous finite valued logic systems[1-3,7], such as, Lukasiewicz system, Post system, Sobocinski system and Godel system, the truth-values domain is finite chain. Recently, there are some discussions about four-valued logic in which the truth-values domain are not chains. Belnap introduced a logic intended to deal in a useful way with inconsistent and incomplete information. This logic is based on a structure called FOUR, which has four truth values: the classical ones, t and f, and two new ones: \perp that intuitively denotes lack of information, and \top that indicates inconsistency. The structure FOUR is a special bilattice. Belnap's logic has no tautologies and hence excluded middle is not valid in it. Arieli and Avron proposed bilattice based logics and corresponding proof systems. Arieli and Avron presented bilattice based four valued logic. In this logic, \perp is the minimal element, \top is the maximal element, t and f are incomparable and the set of designated elements is $D=\{t,T\}$. They also compared four valued formalisms with three valued ones. Yalin Z. and Wenxiu Z. studied the semantics of another four valued logic system K_4^2 , they proved that K_4^2 has no tautologies.

In 1969, Goguen proposed the first lattice-valued logic formal system based on complete-lattice-ordered semigroups where the author does not provide syntax associated with the given semantics. However, the concept of enriched residuated lattice introduced by Goguen provided a new ideal and approach to study the lattice-valued logic. In 1979, based on enriched residuated lattice Pavelka established his famous propositional logic system. Although this logic is based on relatively general lattice, its main results are limited to finite chain or unit interval $[0,1]$ of truth values. In spite of such limitation, these results reflect some fundamental characteristics of lattice-valued logic. In 1982, Novak extended Pavelka's work to first-order lattice-valued logic based on the interval $[0,1]$ or a finite chain, and proved the soundness theorem and completeness theorem of this system. After that, many people still devoted themselves into lattice-valued logic and fuzzy logic in the light of Pavelka and Novak's work.

In order to establish a logic system with truth value in a relatively general lattice, in 1990, during the study of the project "The Study of Abstract Fuzzy Logic" granted by the National Natural Science Foundation of China, Yang Xu etc. firstly established the lattice implication algebra by combining lattice and implication algebra[13], and investigated many of its structure which become the foundation to establish the corresponding logic system from the algebraic viewpoint. Several lattice-valued logic systems[8-13] have been established. As is well-known, it is the difference of degree among attributes of each intermediate-variation-state of things that brings out the uncertainty of things, especially the fuzziness of things. Henceforth, in order to establish the logic system which can efficiently characterize this kind of uncertainty, it obviously needs to gradationally express their difference of degree. So, in [12, 13], dynamical lattice valued logic systems $Lvpl$ and $Lvfl$ are established which can be enough to reflect dynamically gradational technique instead of a static one. Based on these research works, we establish a lattice-valued logic system $lp(X)$ with different axioms in this paper, It may be more convenient in application and study, especially in the case that L is a finite valued lattice. The semantics of $lp(X)$ is discussed. The syntax of $lp(X)$ had been discussed in another paper[15].

2 Preliminaries

Definition 2.1. Let $(L, \vee, \wedge, *, \rightarrow, O, I)$ be a bounded lattice with order-reversing involution " $*$ ", I and O are the greatest and the least element of L respectively, $\rightarrow: L \times L \rightarrow L$ be a mapping. $(L, \vee, \wedge, *, \rightarrow, O, I)$ is called a lattice implication algebra if the following conditions hold for any $x, y, z \in L$:

$$(I1) \quad x \rightarrow (y \rightarrow z) = y \rightarrow (x \rightarrow z)$$

$$(I2) \quad x \rightarrow x = I$$

$$(I3) \quad x \rightarrow y = y^* \rightarrow x^*$$

$$(I4) \quad \text{If } x \rightarrow y = y \rightarrow x = I, \text{ then } x = y$$

$$(I5) \quad (x \rightarrow y) \rightarrow y = (y \rightarrow x) \rightarrow x$$

$$(L1) (x \vee y) \rightarrow z = (x \rightarrow z) \wedge (y \rightarrow z)$$

$$(L2) (x \wedge y) \rightarrow z = (x \rightarrow z) \vee (y \rightarrow z)$$

$(L, \vee, \wedge, *, \rightarrow, O, I)$ is abbreviated to L or $(L, \vee, \wedge, *, \rightarrow)$ for convenience when there is no confusion in this paper.

Example 2.1. Lukasiewicz chain is lattice implication algebra. i.e. $(L, \vee, \wedge, *, \rightarrow, 0, 1)$ is a lattice implication algebra, where

$$L = [0, 1] \text{ or } L = L_n = \{0, \frac{1}{n-1}, \frac{2}{n-1}, \dots, \frac{n-2}{n-1}, 1\}, n=2, 3, \dots$$

$$x \vee y = \max\{x, y\} \quad x \wedge y = \min\{x, y\} \quad x^* = 1 - x \quad x \rightarrow y = \min\{1, 1 - x + y\} \text{ for any } x, y \in L.$$

Example 2.2. $L = C_2 = L_2 = \{0, 1\}$ or $L = L_3 = \{0, \frac{1}{2}, 1\}$ with the same operations as

example 2.1 are also lattice implication algebras.

Theorem 2.1 Let $(L, \vee, \wedge, *, \rightarrow, O, I)$ be a lattice implication algebra. For all $x, y, z \in L$,

- (1) $((x \rightarrow y) \rightarrow y) \rightarrow y = x \rightarrow y$
- (2) $x \vee y = (x \rightarrow y) \rightarrow y$
- (3) $x \wedge y = ((x \rightarrow y) \rightarrow x^*)^*$
- (4) $x \leq y$ iff $x \rightarrow y = I$
- (5) $(x \rightarrow y) \vee (y \rightarrow x) = I$
- (6) $(x \vee y) \rightarrow z = (x \rightarrow z) \wedge (y \rightarrow z)$
 $(x \wedge y) \rightarrow z = (x \rightarrow y) \vee (y \rightarrow z)$
- (7) $x \rightarrow (y \vee z) = (x \rightarrow y) \vee (x \rightarrow z)$
 $x \rightarrow (y \wedge z) = (x \rightarrow y) \wedge (x \rightarrow z)$
- (8) $(x \rightarrow y) \rightarrow (y \rightarrow z) = y \rightarrow (x \vee z) = (z \rightarrow x) \rightarrow (y \rightarrow x)$
 $(z \rightarrow x) \rightarrow (z \rightarrow y) = (x \wedge z) \rightarrow y = (x \rightarrow z) \rightarrow (x \rightarrow y)$
- (9) $x \vee y = I$ iff $x \rightarrow y = y$
 $x \wedge y = O$ iff $x \rightarrow y = x^*$
- (10) $I \rightarrow x = x, x \rightarrow O = x^*, x \rightarrow x = I$
- (11) $x \rightarrow y = O$ iff $x = I$ 并且 $y = O$
- (12) $(x \rightarrow (y \rightarrow z)) \rightarrow ((x \rightarrow y) \rightarrow (x \rightarrow z)) = x \vee y \vee ((x \wedge y) \rightarrow z)$
- (13) If $x \wedge y = O$, then $(x \rightarrow y)^* = (x^* \vee y^*) \rightarrow x = x$
- (14) If $x \vee y = I$, then $(x \rightarrow y)^* = y \rightarrow (x^* \wedge y^*) = y^*$
- (15) $z \leq y \rightarrow x$ iff $y \leq z \rightarrow x$
- (16) $z \rightarrow (y \rightarrow x) \geq (z \rightarrow y) \rightarrow (z \rightarrow x)$
- (17) $(x \rightarrow y) \vee (y \rightarrow x) = I$
- (18) $y \rightarrow z \leq (x \rightarrow y) \rightarrow (x \rightarrow z)$
- (19) $x \rightarrow (y \rightarrow z) = y \rightarrow (x \rightarrow z)$
- (20) $x \rightarrow y \leq (x \vee z) \rightarrow (y \vee z)$

3 Axioms of lp(X)

Definition 3.1. Let X be the non-empty set, $T=\{', \rightarrow\}$ be a type with $ar(\rightarrow) = 2$ and $ar(*) = 1$, the free T algebra on X is denoted by $lp(X)$. The elements of X are called propositional variables, the elements of $lp(X)$ are called formulae.

In $lp(X)$, we define

$$p \vee q = (p \rightarrow q) \rightarrow q$$

$$p \wedge q = (p^* \vee q^*)^*$$

$$p \otimes q = (p \rightarrow q^*)^*$$

Definition 3.2. The following formulae are called axioms of $lp(X)$:

$$A1: p \rightarrow (p^*)^*$$

$$A2: (p^* \rightarrow q^*) \rightarrow (q \rightarrow p)$$

$$A3: p \rightarrow (q \rightarrow p \wedge q)$$

$$A4: p \vee q \rightarrow q \vee p$$

$$A5: (p \rightarrow (q \rightarrow r)) \rightarrow (q \rightarrow (p \rightarrow r))$$

$$A6: (p \rightarrow q) \rightarrow ((q \rightarrow r) \rightarrow (p \rightarrow r))$$

$$A7: ((p \rightarrow r) \wedge (q \rightarrow r)) \rightarrow ((p \vee q) \rightarrow r)$$

$$A8: ((p \rightarrow q) \wedge (p \rightarrow r)) \rightarrow (p \rightarrow (q \wedge r))$$

Denote the class of all lattice implication algebras by LIA. For $L \in LIA$, denote the set of all L -type fuzzy sets on $lp(X)$ by $\mathcal{F}_L(lp(X))$. The symbol O and I denote the lowest and greatest element of L respectively.

Definition 3.3. Let $L \in LIA$, $A \in \mathcal{F}_L(lp(X))$, $p \in lp(X)$. A L -formal proof ω of p from A is a finite sequence of the following form:

$$\omega (p_1, \alpha_1), (p_2, \alpha_2), \dots, (p_n, \alpha_n)$$

where $p_n = p$, for each i , $1 \leq i \leq n$, $p_i \in lp(X)$, $\alpha_i \in L$, and

$$(1) A(p_i) = \alpha_i, \text{ or}$$

$$(2) p_i \text{ is an axiom and } \alpha_i = I, \text{ or}$$

$$(3) \text{ There exist } j, k < i, \text{ such that } p_i = p_k \rightarrow p_j \text{ and } \alpha_i = \alpha_j \otimes \alpha_k.$$

In this definition, n is called the length of ω and denoted by $l(\omega)$, α_n is called the value of ω and denoted by $val(\omega)$. A formal proof of the form:

$$(p_1, \alpha_1), (p_2, \alpha_2), \dots, (p_n, \alpha_n)$$

will be denoted simply by: p_1, p_2, \dots, p_n .

Definition 3.4. Let $L \in LIA$, $A \in \mathcal{F}_L(lp(X))$, $p \in lp(X)$, $\alpha \in L$, p is called an (L, α) theorem of A , denoted by $A \underset{L, \alpha}{\vdash} p$, if

$$\alpha \leq \vee \{ val(\omega) ; \omega \text{ is a } L\text{-formal proof of } p \text{ from } A \}$$

If $A \mid \frac{}{L,I} p$, then it is denoted simply by $A \mid \frac{}{L} p$.

If $\emptyset \mid \frac{}{L,\alpha} p$, we say that p is an (L, α) theorem and write this simply as $\mid \frac{}{L,\alpha} p$.

If $\mid \frac{}{L,I} p$, then p is called a L -theorem and write it simply as $\mid \frac{}{L} p$.

4 Semantics of $lp(X)$

Definition 4.1. Let $L \in LIA$, $A \in \mathcal{F}_L(lp(X))$, $p \in lp(X), \alpha \in L$.

(1) A T-type homomorphism $\gamma: lp(X) \rightarrow L$ is called a L -valuation of $lp(X)$.

(2) p is called a (L, α) tautology of $lp(X)$, denoted by: $\mid \frac{}{L,\alpha} p$, if for each valuation γ of

$lp(X)$, $\gamma(p) \geq \alpha$.

(3) L -valuation of $lp(X)$ γ is said to satisfy A , if $\gamma(p) \geq A(p)$ for each $p \in lp(X)$. A is said to be L -satisfiable, if there exists some L -valuation of $lp(X)$ which satisfy A .

(4) p is called a (L, α) semantical consequence of A , denoted by $\mid \frac{}{L,\alpha} p$, if $\gamma(p) \geq \alpha$ for

each L -valuation γ of $lp(X)$ which satisfy A .

(5) $\mathcal{V}_{L,A,p} = \{ \gamma(p); \gamma \text{ is a } L\text{-valuation of } lp(X) \text{ which satisfy } A \}$

In what follows, the set of all (L, α) tautologies is denoted by $T_\alpha(L)$. When $\alpha = I$ $T_\alpha(L)$ is denoted by $T(L)$.

Theorem 4.1. Suppose $L, L_1 \in LIA$. If L_1 is a lattice implication subalgebra of L then for each $\alpha \in L_1$, $T_\alpha(L) \subseteq T_\alpha(L_1)$.

Theorem 4.2. Suppose $L, L_1 \in LIA$. If $f: L \rightarrow L_1$ is a lattice implication subalgebra isomorphism, then for each $\alpha \in L_1$, $T_\alpha(L) = T_{f(\alpha)}(L_1)$.

By the properties of lattice implication algebras, it is trivial to prove that:

Theorem 4.3. Suppose $L \in LIA$, all the axioms of $lp(X)$ are L -I tautologies.

Theorem 4.4. For each $L \in LIA$, $p, q, r \in lp(X)$, $m, n \in \mathbb{N}$, the following elements of $lp(X)$ are all L -I tautologies.

(1) $p \rightarrow p$ (2) $p \wedge q \rightarrow q$ (3) $p \rightarrow p \vee q$ (4) $q \rightarrow p \vee q$ (5) $p \rightarrow (q \rightarrow p)$

(6) $p \rightarrow (q \rightarrow p \otimes q)$ (7) $(p \rightarrow (q \rightarrow r)) \rightarrow (p \otimes q \rightarrow r)$

(8) $(p^m \rightarrow (q \rightarrow r)) \rightarrow ((p^n \rightarrow q) \rightarrow (p^{m+n} \rightarrow r))$

Proof: We only prove (8), others are easy.

(8) For each L -valuation γ of $lp(X)$, suppose $\gamma(p) = \alpha, \gamma(q) = \beta, \gamma(r) = \theta$. then $\gamma((p^m \rightarrow (q \rightarrow r)) \rightarrow ((p^n \rightarrow q) \rightarrow (p^{m+n} \rightarrow r)))$

$$\begin{aligned}
&= (\alpha^m \rightarrow (\beta \rightarrow \theta)) \rightarrow ((\alpha^n \rightarrow \beta) \rightarrow (\alpha^{m+n} \rightarrow \theta)) \\
&= (\alpha^m \rightarrow (\beta \rightarrow \theta)) \rightarrow ((\alpha^n \rightarrow \beta) \rightarrow (\alpha^m \rightarrow (\alpha^n \rightarrow \theta))) \\
&= (\alpha^m \rightarrow (\beta \rightarrow \theta)) \rightarrow (\alpha^m \rightarrow ((\alpha^n \rightarrow \beta) \rightarrow (\alpha^n \rightarrow \theta))) \\
&\geq (\beta \rightarrow \theta) \rightarrow ((\alpha^n \rightarrow \beta) \rightarrow (\alpha^n \rightarrow \theta)) \\
&\geq (\beta \rightarrow \theta) \rightarrow (\beta \rightarrow \theta) = I
\end{aligned}$$

Theorem 4.5. Let $L \in \text{LIA}$ and L be a complete lattice. For each $p, q, r \in \text{lp}(X)$ the following formulae are L - τ tautologies, where $\tau = \bigwedge_{\beta \in L} (\beta \vee \beta')$.

$$\begin{aligned}
(1) & (p \rightarrow (q \rightarrow r)) \rightarrow ((p \rightarrow q) \rightarrow (p \rightarrow r)) \\
(2) & (p \rightarrow q) \rightarrow ((p \rightarrow q') \rightarrow p')
\end{aligned}$$

Proof: For each L -valuation γ of $\text{lp}(X)$, suppose $\gamma(p) = \alpha, \gamma(q) = \beta, \gamma(r) = \theta$.

$$\begin{aligned}
(1) & \gamma((p \rightarrow (q \rightarrow r)) \rightarrow ((p \rightarrow q) \rightarrow (p \rightarrow r))) \\
&= (\alpha \rightarrow (\beta \rightarrow \theta)) \rightarrow ((\alpha \rightarrow \beta) \rightarrow (\alpha \rightarrow \theta)) \\
&= (\alpha \rightarrow (\beta \rightarrow \theta)) \rightarrow (\alpha \rightarrow ((\alpha \rightarrow \beta) \rightarrow \theta)) \\
&\geq (\beta \rightarrow \theta) \rightarrow ((\alpha \rightarrow \beta) \rightarrow \theta) \\
&= (\alpha \rightarrow \beta) \rightarrow \beta \vee \theta \\
&= \alpha \vee \beta \vee ((\alpha \rightarrow \beta) \rightarrow \theta) \geq \beta
\end{aligned}$$

In the other hand,

$$\begin{aligned}
& \gamma((p \rightarrow (q \rightarrow r)) \rightarrow ((p \rightarrow q) \rightarrow (p \rightarrow r))) \\
&= (\alpha \rightarrow (\beta \rightarrow \theta)) \rightarrow ((\alpha \rightarrow \beta) \rightarrow (\alpha \rightarrow \theta)) \\
&= (\alpha \rightarrow \beta) \rightarrow (\alpha \rightarrow ((\alpha \rightarrow (\beta \rightarrow \theta)) \rightarrow \theta)) \\
&\geq \beta \rightarrow ((\alpha \rightarrow (\beta \rightarrow \theta)) \rightarrow \theta) \\
&= (\alpha \rightarrow (\beta \rightarrow \theta)) \rightarrow (\beta \rightarrow \theta) \\
&= \alpha \vee (\beta \rightarrow \theta) \geq \beta'
\end{aligned}$$

So, it follows that

$$\begin{aligned}
& \gamma((p \rightarrow (q \rightarrow r)) \rightarrow ((p \rightarrow q) \rightarrow (p \rightarrow r))) \geq \beta \vee \beta' \geq \tau \\
(2) & \gamma((p \rightarrow q) \rightarrow ((p \rightarrow q') \rightarrow p')) \\
&= (\alpha \rightarrow \beta) \rightarrow ((\alpha \rightarrow \beta') \rightarrow \alpha') \\
&= (\alpha \rightarrow \beta) \rightarrow (\beta \vee \alpha') \\
&= \alpha \vee \beta \vee ((\alpha \rightarrow \beta) \rightarrow \alpha') \\
&\geq \alpha \vee \beta \vee \alpha' \geq \tau
\end{aligned}$$

Remarks: In classical propositional logic system, these two formulae are tautologies.

Theorem 4.6. Let $L \in \text{LIA}$, $\alpha \in L$, $p, q \in \text{lp}(X)$. Then $\text{Al}_{L, \alpha} \equiv p \rightarrow q$ if and only if

$\alpha \otimes \gamma(p) \leq \gamma(q)$ for each L -valuation γ of $\text{lp}(X)$ which satisfy A..

Proof: For each L -valuation γ of $\text{lp}(X)$ which satisfy A

$\text{Al}_{L, \alpha} \equiv p \rightarrow q$ if and only if $\gamma(p \rightarrow q) \geq \alpha$

If and only if $\alpha \rightarrow (\gamma(p) \rightarrow \gamma(q)) = I$

If and only if $\alpha \otimes \gamma(p) \rightarrow \gamma(q) = I$

If and only if $\alpha \otimes \gamma(p) \leq \gamma(q)$.

Definition 4.2. $A \in \mathcal{F}_L(lp(X))$ is said to be closed if $A(p \rightarrow q) \otimes A(p) \leq A(q)$ for each $p, q \in lp(X)$.

Theorem 4.7. Each L-valuation of $lp(X)$ is closed.

Proof: For each L-valuation γ of $lp(X)$,

$$\begin{aligned} \gamma(p \rightarrow q) \otimes \gamma(p) &= (\gamma(p) \rightarrow \gamma(q)) \otimes \gamma(p) \\ &= ((\gamma(p) \rightarrow \gamma(q)) \rightarrow (\gamma(q))) \vee (\gamma(p) \vee (\gamma(q))) \\ &= \gamma(p) \wedge \gamma(q) \leq \gamma(q) \end{aligned}$$

This completes the proof.

Definition 4.3. Let $p, q \in lp(X), L \in LIA, \alpha \in L$. p is called (L, α) -equivalent with q and denoted by $p \approx_{L, \alpha} q$, if $\vdash_{L, \alpha} p \rightarrow q$ and $\vdash_{L, \alpha} q \rightarrow p$.

Corollary 4.1. For each $p, q \in lp(X), L \in LIA, \alpha \in L$. $p \approx_{L, \alpha} q$ if and only if for each L-valuation γ of $lp(X)$, $\alpha \otimes \gamma(p) \leq \gamma(q)$ and $\alpha \otimes \gamma(q) \leq \gamma(p)$.

We denote $p \approx_{L, I} q$ by $p \approx_L q$. If $p \approx_L q$ for each $L \in LIA$, it will be denoted by $p \approx q$.

Corollary 4.2. For each $p, q, r \in lp(X)$,

- (1) $p \rightarrow (q \rightarrow r) \approx q \rightarrow (p \rightarrow r)$
- (2) $p \rightarrow q \approx q' \rightarrow p'$
- (3) $(p \rightarrow q) \rightarrow q \approx (q \rightarrow p) \rightarrow p$
- (4) $(p \vee q) \rightarrow r \approx (p \rightarrow r) \wedge (q \rightarrow r)$
- (5) $(p \wedge q) \rightarrow r \approx (p \rightarrow r) \vee (q \rightarrow r)$
- (6) $p \rightarrow (q \vee r) \approx (p \rightarrow q) \vee (p \rightarrow r)$
- (7) $p \rightarrow (q \wedge r) \approx (p \rightarrow q) \wedge (p \rightarrow r)$
- (8) $(p \rightarrow q) \rightarrow (r \rightarrow q) \approx r \rightarrow (p \vee q) \approx (q \rightarrow p) \rightarrow (r \rightarrow p)$
- (9) $(p \rightarrow q) \rightarrow (p \rightarrow r) \approx (p \wedge q) \rightarrow r \approx (q \rightarrow p) \rightarrow (q \rightarrow r)$

Lemma 4.1. Let $L \in LIA$, $a, b, c \in L$, if $a = b \rightarrow c$ then $c \geq a \otimes b$.

Proof: $(a \otimes b) \rightarrow c = (a \rightarrow b') \rightarrow c = c' \rightarrow (a \rightarrow b') = a \rightarrow (c' \rightarrow b') = a \rightarrow (b \rightarrow c) = a \rightarrow a = I$.
Therefore $c \geq a \otimes b$.

Theorem 4.8. (Soundness)

1. Let $p \in lp(X), L \in LIA$, if $\vdash_L p$, then $\vdash_{L, \alpha} p$.
2. Let $L \in LIA, A \in \mathcal{F}_L(lp(X)), p \in lp(X), \alpha \in L$, if $\vdash_{L, \alpha} p$, then $\vdash_{L, \alpha} A p$.

Proof: 1. Note that all the axioms of $lp(X)$ are (L, I) -tautologies..

For any L-valuation γ of lp(X), if $\gamma(A)=\gamma(A\rightarrow B)=I$, then

$$\begin{aligned} \gamma(B)=I \rightarrow \gamma(B)=\gamma(A\rightarrow B) \rightarrow \gamma(B) &= (\gamma(A)\rightarrow\gamma(B)) \rightarrow \gamma(B) \\ &= (I\rightarrow\gamma(B)) \rightarrow \gamma(B) = \gamma(B) \rightarrow \gamma(B) = I \end{aligned}$$

So it follows that $\vdash_L p$, then $\vdash_{L,\alpha} p$ for any $p \in lp(X)$, $L \in LIA$.

2. Let each L-valuation γ of lp(X) which satisfy A, p is a (L,α) -theorem of lp(X) from A. We only need to prove: $\gamma(p) \geq \alpha$.

For any (L,α) -formal proof ω from A:

$(p_1, \alpha_1), (p_2, \alpha_2), \dots, (p_n, \alpha_n)$ where $p_n = p, \alpha_n = \alpha$.

(1) If $l(\omega) = 1$, then $A(p_n) = A(p) = \alpha_n = \alpha$ or p_n is a axiom and $\gamma(p_n) = I$.

(i) If $p = p_n$ is a axiom and $\gamma(p) = I$, then it's obvious that

$$\gamma(p) = \gamma(p_n) = val(\omega) = I, \text{ So } \gamma(p) \geq val(\omega) = \alpha.$$

(ii) If $A(p_n) = A(p) = \alpha_n$, from the fact that γ is L-valuation of lp(X) which satisfy A, it follows that

$$\gamma(p) = \gamma(p_n) \geq A(p_n) = A(p) = \alpha_n = val(\omega) = \alpha$$

(2) Suppose that if $l(\omega) < n$ then $\gamma(p) \geq val(\omega)$ for any formal proof ω from A.

Let ω be a formal proof from A with the length n:

$(p_1, \alpha_1), (p_2, \alpha_2), \dots, (p_n, \alpha_n)$ where $p_n = p, \alpha_n = \alpha$.

(i) If $A(p_n) = \alpha_n$ or p_n is a axiom and $\gamma(p_n) = I$, The proof is similar to that in theorem 4.8 1.

(ii) If there exist $i, j < n$ such that $p_i = p_j \rightarrow p_n$ and $\alpha_n = \alpha_i \otimes \alpha_j$, then from the inductive assumption,

$$\gamma(p_i) \geq \alpha_i, \gamma(p_j) \geq \alpha_j, \gamma(p_i) = \gamma(p_j \rightarrow p_n) = \gamma(p_j) \rightarrow \gamma(p_n),$$

From Lemma 4.1 ,it follows that :

$$\gamma(p) = \gamma(p_n) \geq \gamma(p_i) \otimes \gamma(p_j) = \alpha_i \otimes \alpha_j = \alpha_n = val(\omega) = \alpha.$$

So $\gamma(p) \geq val(\omega) = \alpha$ holds for $l(\omega) = n$.

Therefore $\vdash_{L,\alpha} p$, 则 $\vdash_{L,\alpha} p$.

5 Conclusions

In this paper, a lattice-valued logic system lp(X) with different axioms is established. The semantics of lp(X) is discussed. It may be more convenient in application and study, especially in the case that L is a finite valued lattice.

References

- [1] Lukasiewicz, J.: On 3-valued logic. Ruch Filozoficzny 5, 169–170 (1920) (in Polish)
- [2] Post, E.: Introduction to a general theory of elementary propositions. American Journal of Mathematics 43, 163–185 (1921); Reprinted in Van Heijenoort, J.(ed.), From Frege to Godel: A source book in mathematical logic, pp. 1879–1931. Cambridge, mass. (1967)

- [3] Bochar, D.A.: On a 3-valued logic calculus and its application to the analysis of contradictions. *Matematicheskij Sbornik* 4, 287–308 (1939) (in Russian)
- [4] Kleene, S.C.: *Introduction to Metamathematics*. Van Nostrand, New York (1952)
- [5] Zadeh, L.A.: fuzzy sets. *Inf. Control* 8, 338–353 (1965)
- [6] Goguen, J.A.: The logic of inexact concepts. *Synthese* 19, 325–373 (1969)
- [7] Bole, L., Borowik, P.: *Many-valued logics*. Springer, Heidelberg (1992)
- [8] Yang, X., Qin, K.: *Lattice-valued Propositional Logic*. Southwest Jiaotong University 2(1), 123–128 (1993)
- [9] Qin, K., Yang, X.: *Lattice-valued Propositional Logic (II)*. Southwest Jiaotong University 1(2), 22–27 (1994)
- [10] Qin, K., Yang, X., Roh, E.H.: A First Order Lattice Valued Logic System II:Syntax. *The Journal of Fuzzy Mathematics* 4(9), 977–983 (2001)
- [11] Yang, X., Qin, K., Liu, J., Song, Z.: L-valued propositional logic Lvpl. *Int. J. Information Sciences* 114, 205–235 (1999)
- [12] Yang, X., Liu, J., Song, Z., Qin, K.: On semantics of L-valued first-order logic Lvfl. *Int. J. General Systems* 29(1), 53–79 (2000)
- [13] Yang, X., Song, Z., Qin, K., Liu, J.: Syntax of L-valued first-order Logic Lvfl. *Int. Multi. Val. Logic* 7, 213–257 (2001)
- [14] Yang, X., Da, R., Qin, K., Liu, J.: *Lattice-valued logic*. Springer, Heidelberg (2003)
- [15] Li, H.: The Syntax of Lattice-valued Propositional Logic System $lp(x)$. *Journal of Donghua University* (April 2007)

Design of Field Integrative Irrigation Control System Based on Fuzzy Control and PLC

Xiumei Jia, Lili Yao, and Yingmei Zhang

Measuring and Controlling Technology Institute, Taiyuan University of Technology,
Shanxi Taiyuan 030024, China
yaolili-05@163.com

Abstract. In view of the problems that the traditional field irrigation controller can not take into account the impact of air temperature, is poor to drive pumps and inconvenient to be managed and so on, a fuzzy control system for integrative irrigation is designed in the paper. In this design, the power needed is supplied by biogas generator. The biogas digester, generating system, irrigation system, and control system are combined organically. Design mainly improve the traditional three-dimension fuzzy controller, its main control objective is the soil moisture while the secondary control objective is the air temperature; PLC which is stronger at driving is chose as the contrller. The local and remote communications are realized via the Ethernet and GPRS respectively. The system realizes saving water, power and fertilizer, at the same time, realizes monitoring of local and remote which makes it convenient to manage.

Keywords: Integrative irrigation, biogas generation, three-dimensional fuzzy control, PLC, GPRS.

1 Introduction

Nowadays, the extensive agriculture irrigation is not suitable because of the tighter water resource, and there are plenty of resources squandered in rural areas in China[1]. A set of integrative irrigation fuzzy control system saving water, energy and fertilizer for field crops irrigation is designed in this paper in order to build conservation-oriented society and make full use of all useful resources [2].

2 Integral Design

2.1 General Structure of System

In the system, the biogas digester, generating system, irrigation system, control system and monitoring system are combined organically. The block diagram of the general structure of system is shown in Fig.1.

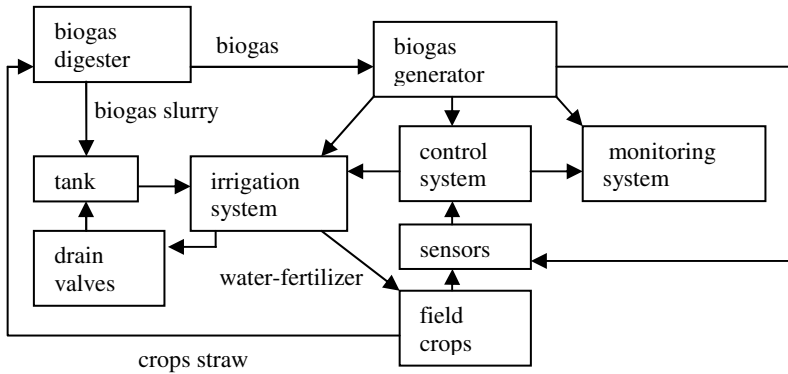


Fig. 1. The block diagram of the general structure of system

2.2 Generating System

Biogas generator vertically inserts a number of exhaust pipes to the biogas digester. The extractive biogas is sent to generator set through the collecting pipe network, and then it is squeezed by the upward piston to be ignited after handled, the burning very-rapidly biogas produces power, so that the crankshaft is driven to rotate, and biogas generator starts to work. A 15KW biogas generator is chosen in the design, it can directly output 12V DC and 220V AC, so it can be used as the main power of PLC, sensors, pump motor and local operating station.

2.3 GPRS Wireless Communication Module

GPRS DTU is selected as the terminal installation for GPRS data transmission in the system, which is embedded TCP/IP protocol and connects with PLC through the RS-232 standard interface [3]. In the data transfer process, the serial data are send to the mobile GPRS network by GPRS DTU at first, then processes the GGSN(GPRS gateway)to the internet web, and then the data are demodulated by the ADSL Modem and forwarded by the router of the data center to the specified database servers after detected by firewall.

2.4 Other Hardware Design in This System

Sprinkling irrigation is used as the irrigation style of the system, Schneider Quantum PLC is used as the controller and 140 CPU 671 60 module is used as the CPU in the system. System uses two sets of PLC, remote I/O and duplicate hot standby system. Sensors, actuators and alarm system are linked to the remote I/O. Temperature

sensors, moisture sensors and solenoid valves are area distribution. There is a drain valve at the end of each branch pipe where is low-lying area for freeze-proofing.

3 Three-Dimension Fuzzy Controller Design

3.1 The Inputs and Outputs of the Fuzzy Controller

Soil moisture and air temperature play a key role on crops growth, while the air temperature is difficult to control in the field environment. When the temperature is relatively high, the using of sprinkling irrigation method can reduce the air temperature above of the farmland. But when temperature is relatively low, the crops stop growing and sprinkling irrigation will not play a significant role on air temperature, especially when air temperature is lower than 0 degree, the water falling on the leaves of crops will freeze and crops may be frost-bitten. Therefore, according to the control scheme that mainly controls soil moisture while air temperature is considered secondarily, a three-dimensional fuzzy controller is designed. The deviation of soil moisture E , the rate of change of moisture EC and the air temperature T are the inputs, the irrigating time U is the output. Moisture's deviation refers to the difference of set value and the actual moisture value.

3.2 The Membership Functions of the Fuzzy Variables

The basic universe is $[-4\%, 4\%]$, the fuzzy universe is $[-4,-3,-2,-1,0,1,2,3,4]$, the quantifying factor is 1, and the fuzzy language variables are $\{NB,NS,ZE,PS,PB\}$ of the deviation of soil moisture; The basic universe is $[-1\%,+1\%]$, the fuzzy universe is $[-4,-3,-2,-1,0,1,2,3,4]$, the quantifying factor is 4, and the fuzzy language variables are $\{NB,NS,ZE,PS,PB\}$ of the rate of change of moisture; The basic universe is $[-15,45]$, the fuzzy universe is $[-15,-5,5,15,25,35,45]$, the quantifying factor is 1, and the fuzzy language variables are $\{NB(\text{frozen}),ZE(\text{cold}),PS(\text{fitness}),PB(\text{hot})\}$ of air temperature; The basic universe is $[0,32]$, the fuzzy universe is $[0,4,8,12,16,20,24,28,32]$, the quantifying factor is 1, and the fuzzy language variables are $\{NB(\text{off}),NS(\text{short time}),ZE(\text{moderate}),PS(\text{long time}),PB(\text{ultra-long time})\}$ of the irrigating time. The profiles of membership functions of inputs and outputs are shown in Fig. 2.

3.3 Fuzzy Control Rules

Based on the experience of experts in the actual irrigation, the rules of fuzzy control are shown in table 1.

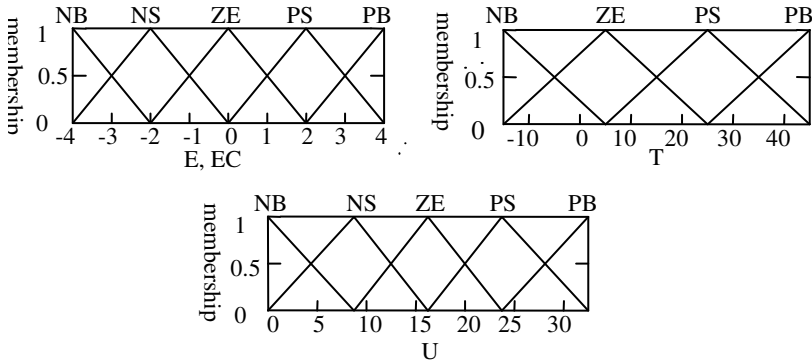


Fig. 2. The profile of membership functions of inputs and outputs

Table 1. The fuzzy control rules

(a) T=NB					
E	EC				
	NB	NS	ZE	PS	PB
NB	NB	NB	NB	NB	NB
NS	NB	NB	NB	NB	NB
ZE	NB	NB	NB	NB	NB
PS	NB	NB	NB	NB	NB
PB	NB	NB	NB	NB	NB

(b) T=NB					
E	EC				
	NB	NS	ZE	PS	PB
NB	NB	NB	NB	NB	NB
NS	NB	NB	NB	NB	NB
ZE	NB	NB	NS	ZE	ZE
PS	NB	NS	ZE	ZE	PS
PB	NB	ZE	ZE	PS	PS

(c) T=NB					
E	EC				
	NB	NS	ZE	PS	PB
NB	NB	NB	NB	NB	NB
NS	NB	NB	NB	NB	NB
ZE	NB	NS	ZE	PS	PS
PS	NS	ZE	PS	PS	PB
PB	ZE	PS	PS	PB	PB

(d) T=NB					
E	EC				
	NB	NS	ZE	PS	PB
NB	NB	NB	NB	NB	NB
NS	NB	NB	NB	NS	ZE
ZE	NS	ZE	PS	PS	PB
PS	ZE	PS	PS	PB	PB
PB	ZE	PS	PB	PB	PB

3.4 Fuzzy Control Tables

The following is the statements' form of fuzzy control used in the design:

If $T = T_i$ and $E = E_i$ and $EC = EC_i$ then $U = U_i$.

The inference system of the fuzzy logic controller in the design is set up by FUZZY that is fuzzy logic GUI tool of MATLAB, and the tables of fuzzy control are calculated by it, too [4]. The wireframe 3-D surface of the fuzzy inference system when $T = PS$ is shown in Fig. 3.

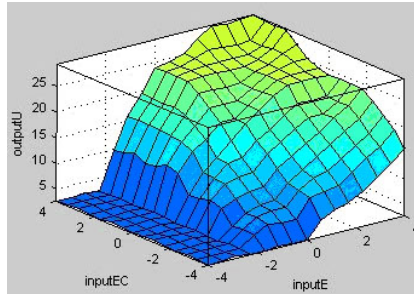


Fig. 3. The wireframe 3-D surface of the fuzzy inference system when $T = PS$

Irrigating time U as the output takes integers between 0 to 32. Taking into account the output of the controller is the opening time of actuators and in order to avoid frequently opening the valves and motors, the values of irrigating time can not be less than 5 minutes, the table of fuzzy control when $T = PS$ is shown in Table 2.

Table 2. The fuzzy control table when $T = PS$

E	EC								
	-4	-3	-2	-1	0	1	2	3	4
-4	0	0	0	0	0	0	0	0	0
-3	0	0	0	0	0	0	0	0	0
-2	0	5	7	8	8	9	9	10	10
-1	0	7	8	10	12	15	15	16	16
0	0	8	10	12	15	20	22	25	25
1	8	12	15	20	23	23	25	25	26
2	9	15	20	23	25	26	26	27	28
3	12	20	23	25	26	26	27	27	32
4	15	20	24	25	26	27	28	32	32

4 Software Design

4.1 PLC Controlling Program

Network configuration, hardware configuration, programming and debugging of PLC are done on the UNITY PRO software. Program is edited by Ladder Diagram. System includes two irrigating modes of manual and automatic. In manual mode, the irrigating time and areas can be set. In automatic mode, PLC reads the temperature and moisture signals measured by temperature sensors and moisture sensors; if the moisture value is within the setting range, it determines whether need to irrigate and the irrigating time according to the fuzzy control rules, otherwise the system runs the deterministic control that the system forcibly opens pump motors and solenoid valves while alarming when the actual moisture value is less than the setting

value. In case of rainy weather or the stop button is pressed, the system stops all irrigation. System automatically opens the drain valves to drain the water in irrigation system after the irrigation has been stopped. The flow chart of main controlling program is shown in Fig. 4.

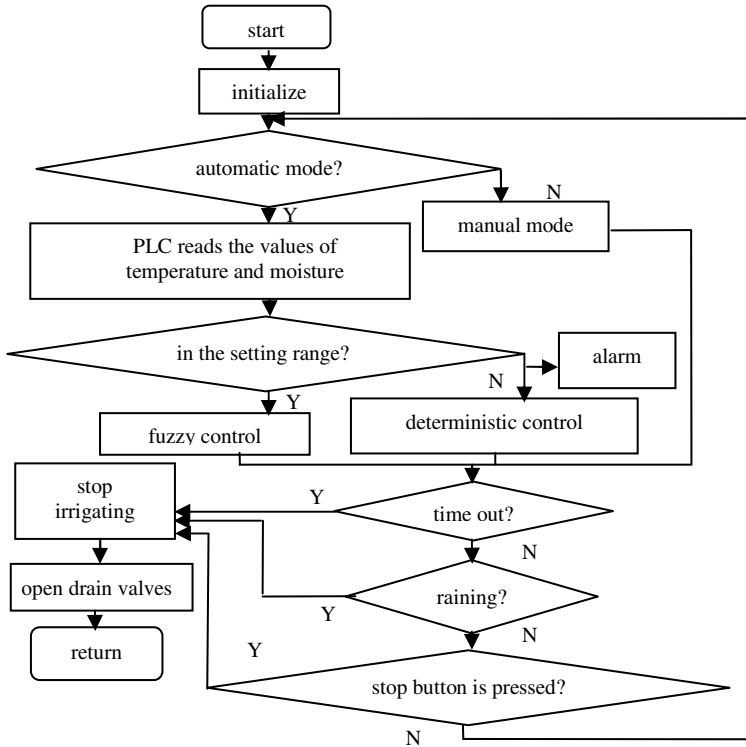


Fig. 4. The flow chart of main controlling program

4.2 Communication Software Design

The communication between local operator station and PLC is carried out through the MODBUS protocol, host computer is the primary station, and PLC is the passive station. The communication between PLC and GPRS module is achieved through AT commands, VC language is used to program. PLC sends corresponding AT commands to complete the initialization of the GPRS module, the setting of the network login, the sending and receiving of the data.

5 Simulation

SIMULINK software of MATLAB is used to model and simulate for the system [5, 6].

The setting value of soil moisture is 70%, after several times of debugging, the curve of soil moisture when $T=PS$ is shown in Fig. 5, in which the sine wave is to be controlled, and the oblique wave has been controlled by the fuzzy controller.

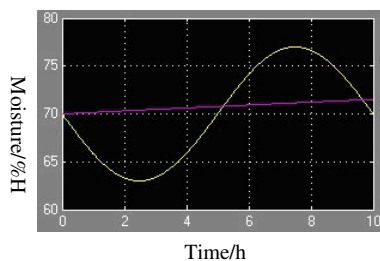


Fig. 5. The simulation curve of soil moisture when $T = PS$

6 Conclusion

A set of integrative irrigation control system based on PLC and fuzzy control for field crops is designed in this paper, which includes the hardware and soft design, the three-dimension fuzzy controller design. The dual control of the soil moisture (main) and the air temperature (secondary), the farmland integration as well as the resources rational utilization are realized. The simulation results meet the requirements of the design standards, which show that the integrative irrigation fuzzy control system responds fast, overshoot slight as well as the control precision of soil moisture is $\pm 4\%$.

References

1. Xie, S., Li, X., Yang, S., He, B.: Design and implementation of fuzzy control for irrigating system with PLC. *Transactions of The Chinese Society of Agricultural Engineering* 23(6), 208–210 (2007)
2. Dong, H., Cui, X., Zhang, N.: Application of Fuzzy Control Technique in an Irrigation Electromagnetic Valve. *Journal of Agricultural Mechanization Research* 2, 150–153 (2008)
3. Chen, L., Feng, J.: Design of Remote Environmental Monitoring System Based on PLC and GPRS. *Automation & Instrumentation* 25(4), 26–28 (2010)
4. Wang, S., Li, J., Shi, S.: The design and application of an automatic irrigation control system based on fuzzy control and wireless wideband communication. *Industrial Instrumentation & Automation* 5, 41–43 (2007)
5. Lu, P., Liu, X.: Design and Emulation on Fuzzy Decoupling Control System of Temperature and Humidity for Greenhouse. *Journal of Agricultural Mechanization Research* 32(1), 44–47 (2010)
6. Shi, X.: *Fuzzy Control and MATLAB Simulation*. Tsinghua University Press, Beijing (2008)

Research on the Information Integration Model Based on Multi-agent

Zhang Yingjie², Ma Quanzhong^{1,3}, Wen Liangjun¹,
Li Guodong¹, and Hu Pengfei¹

¹ School of Control and Computer Engineering, North China Electric
Power University, Beijing, China

² National Science Library, Chinese Academy of Sciences (NSLC), Beijing, China

³ HeNan XuChang Power Supply Company, HeNan, China

zhangyj@wanfangdata.com.cn, maqz@sina.com,
wenlj0530@gmail.com, lgdlgl_bj@yahoo.com.cn, kyofei@126.com

Abstract. This paper proposes a common information integration model based on Multi-Agent and defines the communication protocol for the Agents on the base of the model. Analyze the Agents' communication process and the transmission control mechanism. And it forms a completed Multi-Agent based information integration scheme.

Keywords: Integration model, multi-agent, message set.

1 Introduction

Agent is the concept of artificial intelligence research, the features of adaptability, social and activity it has make it gradually to become an important direction of research of computer science and artificial intelligence[1]. The combination of agent and other research areas is more and more widely, this paper will integrate multi-agent technology and information technology.

The information integration is to obtain information of physical location dispersed, storage structure different, operating environment different in the data source, the information will be integrated into the local data sources by building a platform or model system, and using some method so all users can obtain objective information but do not need cross-system access. Information integration systems not only use existing resources effectively, save costs, and improve the level of sharing of information, solve the problem of information silos.

There are three main methods of information integration Currently, a dedicated integrated interfaces, shared database and integration platforms. We will use the method of shared database, and introduce multi-agent technology to design a multi-agent model based information integration where each data node is similar, and all the nodes share content in the distributed database through the communication mechanism, and the model has the ability to integrate legacy systems.

2 Construction of the Model

There are three main issues to resolved for actual information integration: first is dispersion of the data distribution; second is issues for "legacy" systems; the last is that the structure of the company or organization may be quite large, and also the structure will change, and the interaction of them is very complex.

We will build the models using node as a unit, nodes can be extended infinitely, for this, a multi-node information integration system can be constituted. Formal description of the model can be expressed as:

Integration System =

{(Node_i, Communication Mechanism)|Node_i = Integration System or Node System Model}

The architecture for a single integrated node of our model is shown in Fig 1.

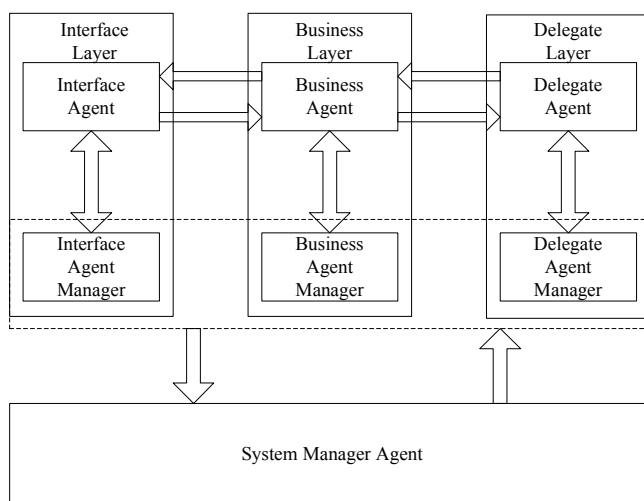


Fig. 1. The architecture of single node

The roles of the nodes described as follows:

1. Delegate Agent and Delegate Agent Manager.
2. Business Agent and Business Agent Manager.
3. Interface Agent and Interface Agent Manager.
4. System Manager Agent.

The node can be divided into interface layer, business layer and agent layer not only vertical structure and process structure. Each agent is the basic unit of activity entities (known as the worker agent), while each layer has a agent manager (called manager agent) that is responsible for activities of the layer to interact with other layers [2].

The features of Systems Management Coordination Agent (SMA): it is the core of the integrated node that manager the agent manager for each layer, and coordinate

communication between agent managers to ensure that the system has no problems, but not to manage specific unit. node directly.

The features of Interface Agent (IA): it provides all the access interfaces for user or other client and needs to interact with specific business functions.

The features of Interface Agent Manager (IAM) :it manages the IA, coordinates IA and user requests, intervene the request of IA if necessary, or distributes the IA request to the specific business agent. As the manager of integrated node, it is responsible of SMA and interacts with the SMA.

The features of Business Agent (BA): as business processing modules, each of the different business functions will be handled by the specific agent, BA implement the task which interface layer sends and all the BAs are managed by the BAM.

The features of Business Agent Manager (BAM): it manages the BA and coordinates them which configurations some scheduling algorithm and coordination mechanism. The BAM receives information requests from IA, distributes tasks in accordance with the algorithm and mechanism, collects BA processing results back to the interface layer.

The features of Delegate Agent (DA): there are two functions, one of them is the data processing as a data agent; the other is responsible for data exchange with other nodes as request broker. The DA will send requests to other nodes when the node itself can not complete the requests from users.

The features of Delegate Agent Manager (DAM): it manages and coordinates the work of all DAs, the management of the node to other nodes in DA's request and distribution, and interacts with the SMA to exchange the status information between DA and DAM.

A complex information integration model can be described as shown in Fig 2 in summary.

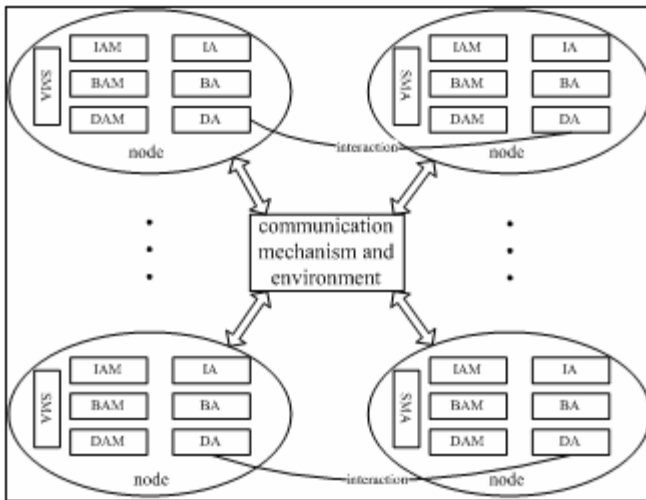


Fig. 2. The model of data integration system

3 Communication Mechanism

The message set about information integration of node system is the basis for communication between agents. The development of message set involves following aspects: the message format, message type and name, the type of messages which each agent can receive, and their action after receiving the message.

3.1 Format of Message

The message set about information integration of node system is the basis for communication between agents. The development of message set involves following aspects: the message format, message type and name, the type of messages which each agent can receive, and their action after receiving the message.

The message structure includes the following in node system: the message name, message type (equivalent to the type of message), the message sender, message recipient, the specified reply object, the session ID and other fields[3], and its parametric representation of the fields as follows:

(node system message parameters:

```
: sender
: receiver
: feedback-to
: mname
: mtype
:content
:conversation_id
:timestamp
:extension
)
```

The meaning of each field as follows:

Sender: the sender of the message

Receiver: Receiver of the message

Feedback-to: The specified message object replied. It is that when the receiver receives the message and generate a corresponding feedback message after handling. It is necessary that the sender specify the receiver object of the feedback message in some cases.

Mname: It is the abbreviation of message name, such as the interaction message is "Interact Message", the name of data query request message is the "Date Search Request Message".

Mtype: It is the abbreviation of message type or action type, the message purpose of sender can be defined according to message type. Fipa-acl standard is used in this field[3], the following messages of message set with the corresponding the fipa-acl standard.

Content: Content of message

Conversation_id: Session ID, it is used to identify a session, the session consists of multiple, independent message about two agent all these messages have the same Conversation_id.

Timestamp: It is used to record the time of the node system when the message is sent. A message can be uniquely identified by the session ID and time stamp through the message sender.

Extension: Message extension field, which can contain a number of special fields. The message structure above is general message structure in the node system, different types messages will be exist, there will be some special fields, but these fields are not shared by all messages, this type field will appear in the extension.

3.2 Message Set

On the one hand, the message can be divided into four categories according to the role in the message set of node system (node message set for short), first is interactive messages, worker agent send requests to IAM before requesting collaborative work with other types of worker agent, these request is done by IAM. After receiving the request, IAM send it to corresponding worker agent periodically. Second is feedback messages, including confirmation messages and feedback messages of between IAM and worker agent. Third is the interaction and cooperative message between the worker agents, worker agents collaborate with each other to complete tasks. Fourth is the management and collaboration message between manager agent and the worker agent, including the management, support and cooperation manager agent to worker agent.

On the other hand, there are three type message according to the source where the message send, they are message of interface layer, message of business layer and message delegate layer with analyzing the structure of the node system.

Table 1 shows message name of the node system and the relationship between roles.

Table 1. The message name and role relationship

Message Name	Sender	Receiver
IA Interact Message	IA	IAM
Data Request Message	IA	IAM
BA Interact Message	BA	BAM
BA Allocated Report Message	BAM	BA
Data Request Feedback Message	BA	IA
DA Allocated Report Message	DAM	DA
Node Data Request Message	DA	DAM
Data Request Feedback Message	DAM	DA

4 Design and Implement

According to the former integrated model, the information integration system can be designed, built as Fig 3 shows.

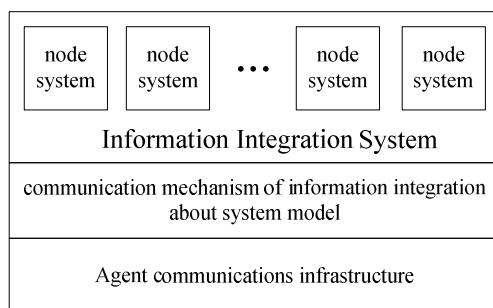


Fig. 3. The system architecture diagram

With refining the system nodes, the task decomposition of each node shows in Fig 4. As the core function about information integration, there are local collection, remote information acquisition and integration of legacy systems three methods, in addition to user interaction, coordination and information organizations as supporting functions.

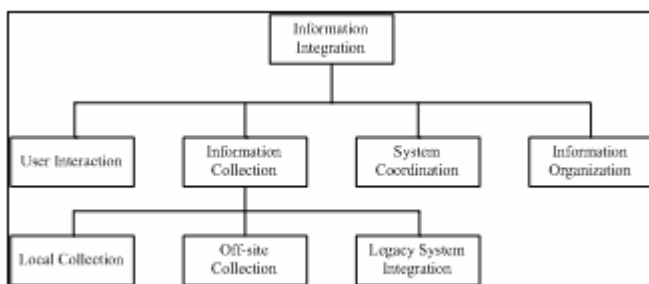


Fig. 4. The task decomposition on data nodes

The function description of each task module as follows:

User interaction: it is responsible for interacting with the user, providing interactive interface, approaching to user needs, explaining and classifying users' needs according to knowledge, and form the format which of the system can understand, generate the task system need, also is the interface that system and user interact with each other.

Information Collection: Data from three sources, first is local database, in which each data node has a database which structure is same, the agent that does data collection first query the database which is cached in local node. second is that it can access data from other database of node by doing request to other data node. Third is that information integration through the legacy systems output.

System Coordination: The agent in system has many roles, there should be monitoring and coordination mechanisms, monitoring actions of members supported, and is responsible for the user's login and registration management, life cycle management of system members.

Information Organization: It is responsible for that the data format conversion of information collected, write to the local database, And do some evaluation and adjustment of knowledge of the information in the database.

There are two types of roles in node system: worker agent and manager agent; worker agent is responsible for completing specific tasks, while manager agent is responsible for the creation of the former, communicate tasks, coordinate and manage the activities of worker agent; In addition, worker agent can communicate with each other, However, they are managed and coordinated by manager agent, after completing the task, the manager agent decide to assign its fate (destruction or returned to the buffer pool).

According to information integration model based on multi-agent described above, the system can be divided into some modules by function, after organizing and assigning, a simplified class diagram describing the prototype system shown in Fig 5.

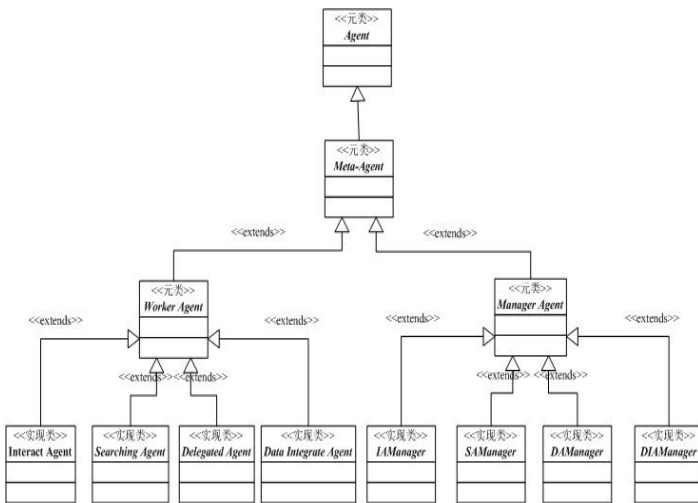


Fig. 5. Agent class model of node system

5 Conclusion

This paper introduces multi-agent technology into the field of information integration, the purpose is to create an open integrated solution for general information based multi-agent with agent's own initiative, adaptability and other characteristics. Of course, this paper only complete the model based on the ideas above. More multi-agent platform should be run for application of the node system, and constantly improve the model, the ideal situation is that it is able to establish information integration platform environment based on multi-agent finally that can automatically provide specific application solutions who need to build such as enterprise or government and other organizations.

References

1. Wooldridge, M., Jennings, N.R.: Intelligent agents: theory and practice. *The Knowledge Engineering Review* 10(2), 115–152 (1995)
2. Pitt, J., Mamdani, A.: Designing agent communication languages for multi-agent system. In: Garijo, F.J., Boman, M. (eds.) *MAAMAW 1999*. LNCS, vol. 1647, pp. 102–114. Springer, Heidelberg (1999)
3. FIPA Communicative Act Library Specification, SC00037 (December 3, 2002)
4. Nickles, M., Rovatsos, M., Weiss, G.: Agents and computational autonomy: potential, risks, and solutions, pp. 80–81. Springer, Heidelberg (2004)
5. Cohen, P.R., Perrault, C.R.: Elements of a plan based theory of speech acts. *Cognitive Science* 3, 177–212 (1979)
6. Gruber, T.R.: Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human Computer Studies* 43(5/6), 907–928 (1995)

Hardware Design of T/Q/IQ/IT for Intra 16×16 in H.264/AVC

Meihua Gu and Jinyan Hao

College of Electronic Information, Xi'an Polytechnic University,
Xi'an, China
gu_meihua@163.com, jinyan_554@126.com

Abstract. This paper presents an efficient hardware architecture design for a dedicated transforms and quantization loop for intra16×16. Reusable hardware structure, and parallel processing in pipeline are adopted for DCT and Hadamard implementation. In order to enhance the speed of image processing, quantization, inverse quantization module also uses the parallel and pipeline structure. The design is described by Verilog language, implemented with 0.13μm CMOS technology. Experimental results show that the processing speed of the proposed architecture can reach 548 fps for the video resolution of 1920×1080, which is far beyond the requirements of real-time processing requirements.

Keywords: H.264, intra16×16, T/Q/IQ/IT, hardware design.

1 Introduction

H.264/AVC [1] is the latest video coding standard. Compared with MPEG-4, H.263, and MPEG-2, H.264/AVC can achieve 39%, 49%, and 64% of bit-rate reduction, respectively [2]. Intra Prediction plays an important role in H.264 coding flow. The Intra Prediction module explores the spatial redundancy inside a frame, exploring similarities of neighbor blocks. This is made using the data of one previously processed block to predict the current block [3]. In the coding process, the H.264/AVC standard defines that the residues between the original and the intra-predicted block must pass through the Forward Transforms (T), Forward Quantization (Q) Inverse Quantization (IQ) and Inverse Transforms (IT) to generate the reconstructed blocks that will be used as references by the Intra Prediction to predict the next block [3]. Thus, the Intra Prediction module must be idle, waiting for the reconstructed block while this block is being processed by the T/Q/IQ/IT loop. Then, the number of cycles that a block needs to be processed by the T, Q, IQ and IT modules affects the performance of the Intra Prediction. This means that the latency of the architectures designed for this reconstruction path must be as low as possible and the throughput of these modules must be as high as possible.

In this paper we propose a hardware design of TQ/ITIQ for intra16×16 prediction functions in real-time H.264/AVC encoding. The rest of this paper is organized as following. In Section 2 we describe intra 16×16 transform flow. Section 3 describes

transform and quantization algorithm. In Section 4 we present our hardware design. In section 5 we show our experimental result. Finally we draw a conclusion in Section 6.

2 Intra16x16 Transform Flow

For residual 16x16 data block in the frame, the first is that we can divided it into small pieces of 16 4x4, we can respectively process DCT transform for 16 4x4 small block according to the order shown in the Fig.1, after we can again process the 4x4 Hadamard transform for DC coefficients of each 4x4 block .Then we can process the quantization, reverse quantization in turn for each 4x4 block of change coding, after reverse quantization for 4x4 block composed of DC coefficient , first of reverse Hadamard change, we can put each DC coefficient in the corresponding each 4x4 block, then reverse DCT transform for each 4x4 block , the 4x4 block of DCT transform add with the data of prediction pieces which position is corresponding to 4x4 block, we get the data of reconstruction pieces. The flow chart about detailed algorithm of 16x16 transform coding in the frame as shown in Fig.2.

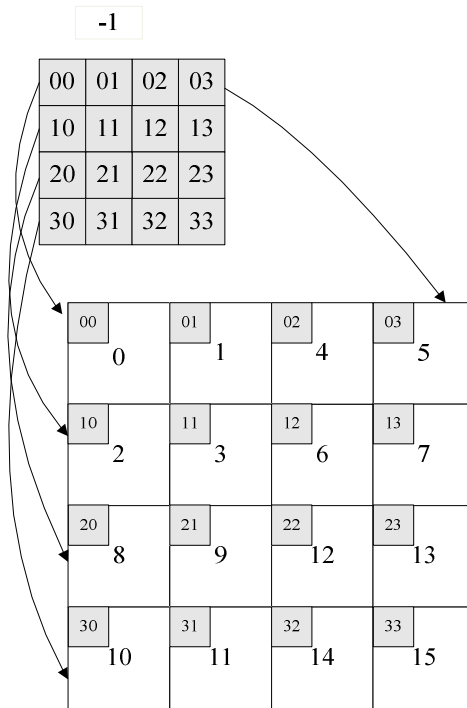


Fig. 1. Scanning sequence for residual data

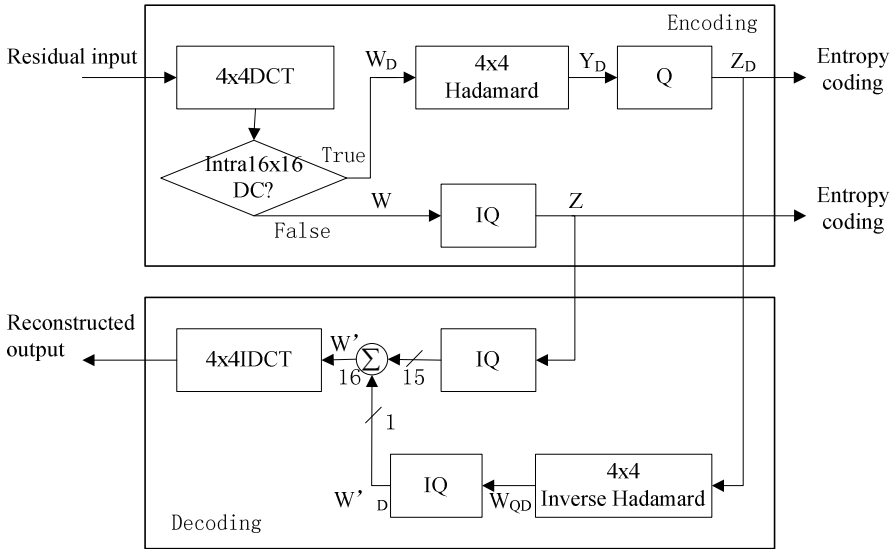


Fig. 2. Intra16x16 Transform flow only for luma samples

The below is detailed instructions for Intra16x16 circuit design. Circuit design chooses synchronous timing, synchronous trigger, synchronous reset. Transform coding is the larger module of computation in the whole h.264 coding system, the influence of its speed throughout the speed of coding system is also larger. In order to improve the circuit speed, DCT transform, quantification, inverse quantification, inverse DCT transform adopt parallel lines structure.

3 Transform and Quantization in H.264

H.264 supports a hierarchical transform path which includes two levels of different transforms. The first transform level is a 4x4 integer DCT-based transform (also called ‘core’ transform) that operates on the input 4x4 residual blocks. The transform operation can be written as (1):

$$W = C_f X C_f^T \tag{1}$$

Where the matrix X is the input 4x4 residual block, and C_f is given by the following

$$C_f = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix} \tag{2}$$

The second level of transform (Hadamard transform) is used in two cases with two different sizes and is applied only on the DC coefficients gathered from each transformed residual block in the first stage.

For the case of a macroblock coded in 16x16 Intra prediction mode (also known as ‘16Intra’), its luma pixels are first transformed using the ‘core’ transform described earlier and, as a second step, a gathered 4x4 DC coefficients block is transformed again using a 4x4 Hadamard transform as described by the following set of equations

$$Y_D = \begin{pmatrix} [1 & 1 & 1 & 1] \\ [1 & 1 & -1 & -1] \\ [1 & -1 & -1 & 1] \\ [1 & -1 & 1 & -1] \end{pmatrix} W_D \begin{pmatrix} [1 & 1 & 1 & 1] \\ [1 & 1 & -1 & -1] \\ [1 & -1 & -1 & 1] \\ [1 & -1 & 1 & -1] \end{pmatrix} / 2 \tag{3}$$

Where W_D is 4x4 DC parameters, Y_D is the result of Hadamard transform.

It should be noted that luma residual blocks predicted in prediction mode other than 16x16 intra prediction mode are transformed with core transform only.

H.264 utilizes a scalar quantization formula that expresses all scaling factors in fixed point arithmetic, in a way to substitute integer division by an integer multiplication and a shift. This formula can be directly applied to the output coefficients, W_{ij} (for the core transform) and $Y_D(i,j)$ (for the Hadamard transform), of the transformation process. The following equations show the quantization formulas used in these two cases.

$$Z_{ij} = round(W_{ij} \cdot \frac{MF_{ij}}{2^{qbits}}) \tag{4}$$

$$Z_{D(i,j)} = round(Y_{D(i,j)} \cdot \frac{MF_{00}}{2^{qbits+1}}) \tag{5}$$

$$qbits = 15 + floor(QP / 6) \tag{6}$$

QP is the integer quantization parameter that takes a value in the range [0-51] and MF is a multiplication factor that takes positive integer values, depending on QP and the element’s position in the matrix as stated in article[4] and [5].

In the above and subsequent formula, “>>” indicates a binary shift right, and f is $2^{qbits} / 3$ for Intra predicted blocks or $2^{qbits} / 6$ for Inter predicted blocks.

4 Hardware Design

4.1 Transform/Inverse Transform Circuit

According to rapid wing algorithm, the decomposition of coefficient matrix W can reduce operation times. The matrix representation form from Hadamard transform and DCT transform can be observed, of which are only a few factors and their decomposition

form is basically similar, Fig.3 is fast papilionaceous algorithm flowchart which is corresponding to DCT/Hadamard transform formula, it can be seen from the drawing , the differ is only a multiplication factor, and Hadamard transform and DCT transform is time-sharing in time, so we can design a selector in DCT transform unit, and control the selection of multiplication factor using sequential control signals according to need.

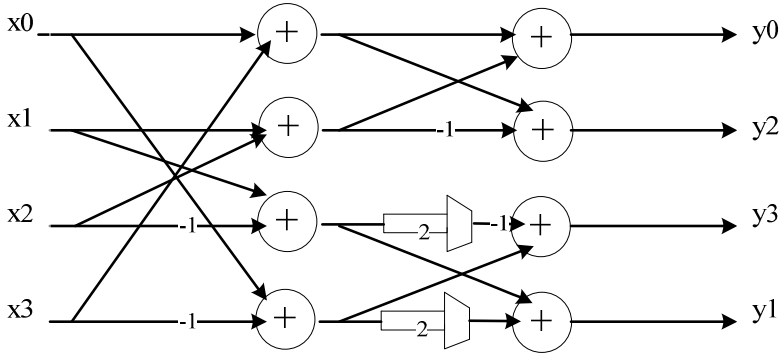


Fig. 3. 1D-DCT /Hadamard fast papilionaceous flowchart

Fig.4 is fast papilionaceous algorithm flowchart which is corresponding to IDCT/IHadamard transform formula, it can be seen from the drawing, the differ is only a multiplication factor, and IDCT transform and IHadamard transform is time-sharing in time, so we can design a selector in IDCT transform unit, and control the selection of multiplication factor using sequential control signals according to need, so the methods of circuit design is similar to the methods of DCT transform.

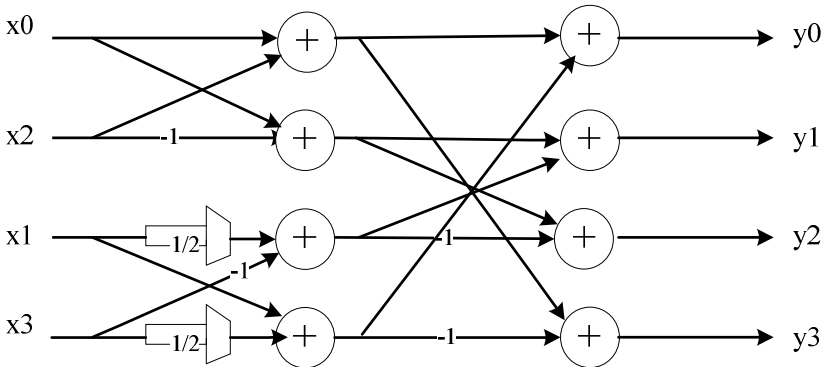


Fig. 4. 1D-IDCT /IHadamard fast papilionaceous flowchart

For formula $Y = WXW^T$, we can transform for $Y = W(WX^T)^T$, the design of hardware circuit for DCT transform code fulfill by transform the second matrix

transformation into two one-dimensional matrix transformation. Fig.5 is hardware structure chart of DCT transform, with the pipeline architecture structure. For a 4x4 block, there are 16 input data ,which input into the circuit by the parallel manner, every PE unit process 4 data of a line one-time , need four such PE units, the results of the first level output meanwhile ; output results of the first level were transposed, as the input of second level according to the list of ways, the second 1D_DCT and the first level 1D_DCT choose the same fast algorithm, this part of the circuit structure and the first part have the same PE unit.

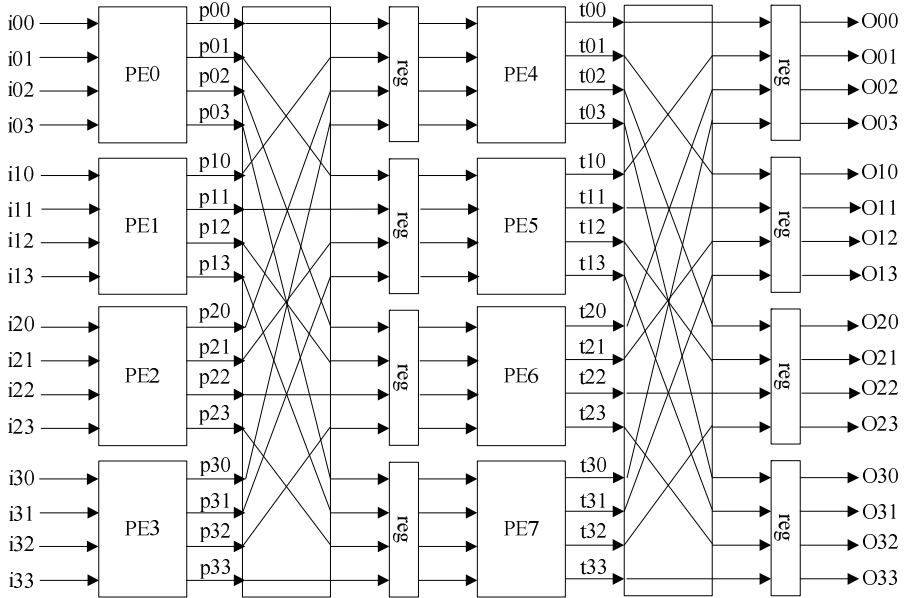


Fig. 5. 2D-DCT hardware architecture

4.2 Quantitative/Inverse Quantitative Circuit

Quantization multiplicities and shifts the so efficient which are after transform coding, so as to compress data further. Coefficient is currently quantitative coefficient, judge input data is DC coefficient or AC coefficient control signal, x_position and y_position says the position which the current to quantitative coefficient is in block's, and final output is quantified data; This circuit consists of multiplier, Qpmod unit, shift unit and Qpdiv unit four sub-modules. The PE unit structure of quantitative circuit diagram is shown in figure6 shows, the whole Intra16x16 adopt line structure, and deal with 16 data.

When $0 \leq QP \leq 5$, for different QP, its corresponding scale factor is different. With 6 and QP take the module. If results are consistent, they have the quantification factors. Transform coefficient and quantification factors' multiplication needs to shift. The specific digits are related to the remainder of (displacement QP / 6), So the role of div6 is choose different shift factor according to different QP.

The aim of inverse quantification is to put the conversion coefficient after quantization restore, so that inverse transformation can rebuild the data of pictures. Its inverse transform circuit structure and quantification are similar, just the scaling factors stored are different. We no longer detail here. Scaling factors of inverse transform can also be referenced in article[5].

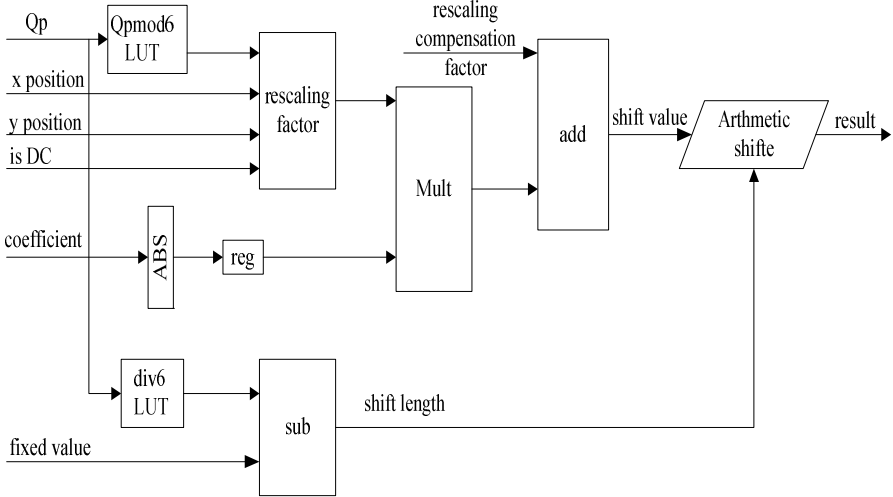


Fig. 6. Calculating circuits and PE architecture

5 Experimental Result

The Proposed structure is described in Verilog Language, and simulated in VCS. Test files are written according to the actual need, results show that the simulation results of DCT module, quantization module, inverse quantization module, and IDCT module are consistent with MATLAB calculation results respectively.

Fig.7 shows the simulation results of the whole DCT/Q/IQ/IDCT module. Key signal is that end_Intra is end enabled signal, for rec_00, rec_01,... rec_77 is the image information after the reconstruction, we can see that VCS simulation results are same to Matlab fitting results.

Described by the Verilog-HDL, verified under the VCS environment, after by being compiled and simlinked, it is carried on the comprehensive by using process repository of SMIC 0.13 μ m and applying the design compiler of Synopsys company. Maximum path delay is 2.75826 ns with a total area of 1127015.885466 μ m². For HD video sequence with sampling format for 4:2:0 and resolution for the 1920 \times 1080 , its processing speed can reach 548.7fps, is far beyond 30fps of real-time processing requirements.

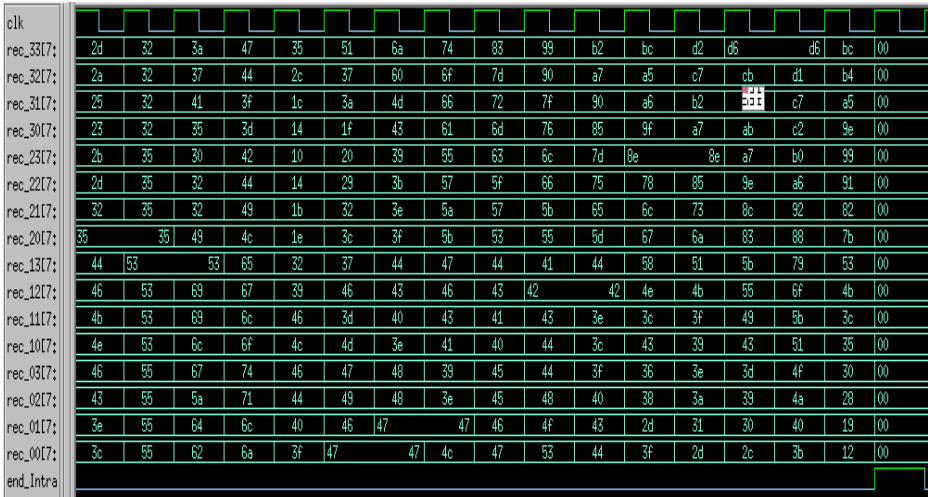


Fig. 7. Simulation waveform of final

6 Conclusion

This paper proposed an effective transform, quantization, and inverse transform, inverse quantization loop module for intra 16x16 in H.264. 16 4x4 integer DCT transforms and 1 4x4 Hadamard transform share the same architecture to save hardware resources. Quantization and inverse quantization also has the similar architecture except the different scaling factors and with no shifts. Experiments show that the proposed architecture can satisfy the real time processing requirements for 1080p video.

Acknowledgments. This paper was supported by Scientific Research Program of Shaanxi Province Department of Education (2010JK558).

References

1. Joint Video Team, Draft ITU-T Recommendation and Final Draft Inter-national Standard of Joint Video Specification, ITU-T Rec. H.264 and ISO/IEC 14496-10 AVC, Geneva (2003)
2. Joch, A., Kossentini, F., Schwarz, H.: Performance Comparison of Video Coding Standards Using Lagragian Coder Control. In: IEEE International Conference of Image Processing, pp. 501–504. IEEE Press, New York (2002)
3. Wiegand, T., Sullivan, G.J.: Overview of the H.264/AVC Video Coding Standard. IEEE Transactions on Circuits and Systems for Video Technology 13(7), 560–576 (2003)
4. Malvar, H.S., Hallapuro, A., Karczewicz, M.: Low Complexity Transform and Quantization in H.264/AVC. IEEE Transactions on Circuits and Systems for Video Technology 13(7), 598–603 (2003)
5. H.264 / MPEG-4 Part 10: Transform & Quantization, <http://www.vcodex.com>

Construction of Multi-Agent System for Decision Support of Online Shopping

Jiao Li and Yuqiang Feng

School of Management, Harbin Institute of Technology
Harbin Heilongjiang, P.R. China, 150001
mlbbconan@163.com

Abstract. In recent years, both the number of online E-commerce sites and the types of goods have become more and more. However, because of it, customers have to integrate and analyze much more goods information to make shopping decision during online shopping. This study introduced a novel multi-agent model to automatically gather and provide products' information to support the customer's decision of online shopping. In this study, the user preference is applied to product searching process and XML Schema is used to gather cross-site data to achieve a personalized product recommendation. This study provides a forward-looking view for E-commerce intelligent decision support system research.

Keywords: Multi-Agent System, XML Schema, personalized recommendation, online shopping decision support.

1 Introduction

According to the 26th China Internet Development Statistics Report of China Internet Information Center, up to June 2010, the scale of users shopping online has reach 142 million and the utilization of online shopping has raised to 33.8%, going up by 5.7 percentage points. In these six months, the total of user increased by 31.4%. The scale of users shopping online is increasing rapidly, which shows strong development momentum of E-commerce market in China [1].

The growth of E-commerce provides infinite convenience for customers, as well as brings a wealth of optional product range and detailed and a huge amount of product information. As a result of the constraints of cognitive ability, time and online shopping experience, customers' ability of collecting and analyzing relevant information is limited when they are shopping online. In face of so huge information, it's becoming harder and harder for customer to sift the information they need and make shopping decision according to these information[2]. In order to help customer to obtain important commodity information as soon as possible and make decision in the light of comprehensive information, this paper is aimed at researching Multi-Agent System for decision support of online shopping. In this paper, we search product information amount the entire network by introducing user preference and personalized recommendation, and then give the architecture and implement method of this system.

Agent technology originated from Artificial Intelligence in 60s last century. In 80s, because of the increase of the level of computer hardware and software and the extensive demand of Artificial Intelligence, agent technology has developed significantly. As a software object in specific operating environment, agent technology has situational, autonomy and adaptability features [3]. Due to these features, agent technology has unique advantage in business decision-making, semi-automated information acquisition and analysis. However, human intelligence not just exists in the individual thinking activities, also in co-operation among groups of human society. So researching Multi-Agent System (MAS) can better embody human society intelligence, and better suit open and dynamic world environment. Moreover, agents have to deal with much uncertain, even contradictory information, which is very difficult for single agent system. So it's necessary to research Multi-Agent System.

The research of MAS focuses on how to cooperate the behavior of a group of agents, that is how to cooperate their knowledge, goal, strategy and plan etc. in order to solve the problem jointly. In recent years, scholars from different fields have done a lot of further researches in this area. Van Katwijk, R.T. [4] researches controlling traffic signal network using MAS. Yan Cao [5] researches multi-layer intelligent encapsulation of manufacturing resources for MAS-based distributed manufacturing system. Ying Zhang [6] introduces how to model and develop MAS applying Process Configuration Pattern and how to make a software system. Zhichao Ren [7] applies lightweight object management framework, Spring, which is the most common in Java system, develop main framework of MAS-based mobile positioning system, AGPS. In conclusion, we have the technical foundation to achieve the system, which is feasible.

2 Design of System Functions and Structure

2.1 Design of System Functions

The purpose of this study is to design a Multi-Agent system for decision support of online shopping. When customers tend to shop online, the system could automatically search the suited products through massive shopping information and provide personalized service. To achieve the purpose, the system needs to realize the following functions:

- It is based on the customer's requirements and past search history, to determine the user's buying tendency and automatically search the related products information (including prices, recent trading volume, the credibility of the seller, the buyer reviews, etc.) in available e-commerce sites.
- Combined with the products information database which stores the information which products the customers having the same search request finally click, to make personalized recommendation of products.
- According to the attributes that customers value, the system will compare and rank the search results and show the list to the customers to support shopping decision.
- To store the list of searching products into database and form the search history.

- Each e-commerce site has its unique structure, the search results of the vocabulary and notation, so it is necessary to dynamically identify this structural difference to achieve data capture cross-platform.

2.2 Design of System Workflow and Structure

As a personal customized service, the system provides customers with personalized shopping guide features and maintains preference information for each registered customers. When a customer registers, the system will initialize a default preference. And in the after use process of registered customers, the real preferences changing will reflect in the actual purchase behaviors. The system will constantly adjust its preference knowledge to the real preferences of the customers.

The main workflow of the system is: customers submit a list of feature information of the products they want, then the system search in the database and internet. For the search results, the system will evaluate and filter them based on the customers' preferences and constantly show the search progress and the current selected products to the customers. Customers could see current search progress and intermediate results of the search. After customers select the products they interest in, the system will adjust its preference knowledge to the real preferences of the customers, so that the results of the evaluation and selection could best meet their needs.

Within the system, the modules are divided by function and integrated into each agent. And there are three agents in the system: user interaction agent, search agent and information conversion agent. Each agent has its own function, yet cooperates with others to accomplish the task of supporting shopping decision.

This division of module mainly considers functional cohesion and increasing the parallelism. Functional cohesion means that the functions of each module are relatively independent, and increasing the parallelism means each module could run concurrently to improve system performance, there is an inherent relationship between the two techniques. In fact, MAS originally developed from distributed artificial Intelligence. And to solve problems through the division of labor and running concurrently, it is just an advantage of MAS[8]. The division of labor of the agents in the system is that:

User interaction agent: When a customer enters the product keyword, the user interaction agent will capture its historic search preference in the preference database, and forecast its current search preference, then capture the websites to be search in XML Schema database and transmit these information to the search agent for the next search work. Meanwhile, it will search in the products information database to capture suitable products' information and show the result list to the customer ranked by products important attributes.

Search agent: After it receives information from the user interaction agent, the search agent will connect the related e-commerce websites to open web pages according to the search requirements, and transmit the HTML pages to the information conversion agent.

Information conversion agent: When it receives the HTML pages from the search agent, the information conversion agent will extracts HTML pages information according to XML Schema, and form XML documents. Then the products information is to be stored into database in a unified form.

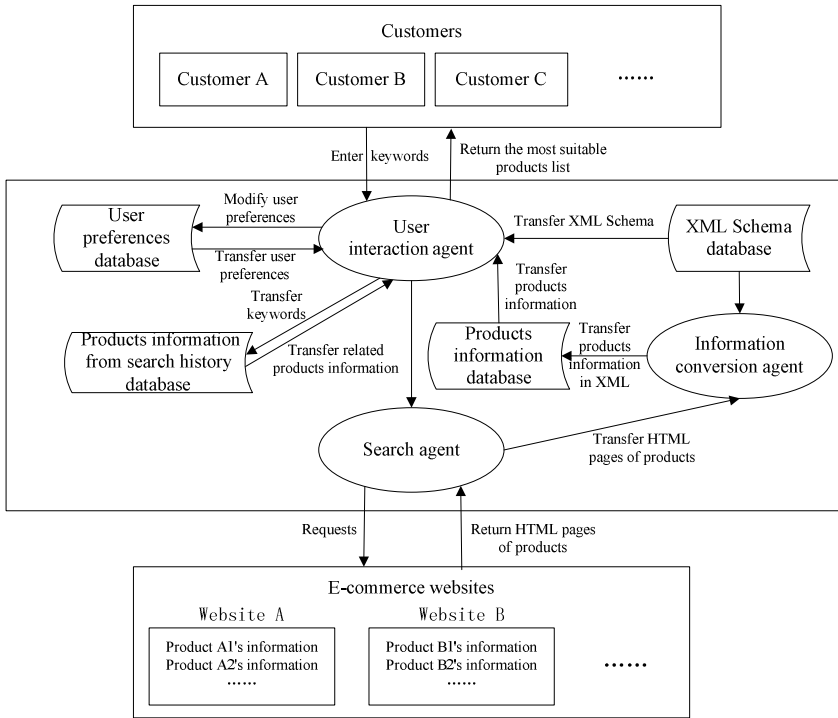


Fig. 1. Structure of Multi-Agent System for decision support of online shopping

3 Analysis of the Value and Technical Points of the System

3.1 Analysis of the Value of the System

At present, e-commerce websites in China, such as TaoBao, PaiPai and JiongDong, have already shown the search result list to customers ranked by default rules and provided ranking by price, sales volume and credit. However, none of these e-commerce websites take customers' preferences into consideration, so customers could not get personalized services. Meanwhile, according to the situation that each website operates independently, the search results can only contain products in its own website and none of the websites can make recommendation of whole network, so it is possible that the most suitable products for customers can not be displayed. Considering present situation and deficiency, this study presented the concept of Multi-Agent system for decision support of online shopping, took customers' preferences and personalized recommendation into consideration, and searched products information in whole network, to provide customers with the most suitable products to their requirements and offer better services.

3.2 Analysis of Technical Points of the System

Firstly, extracting data from HTML to XML: Generally speaking, there exist different forms of return results among each e-commerce website, for example, some sellers expressed the price as `<Product_Price>$45</Product_Price>`, while others expressed it as `<Price>$30</Price>`. However, although each website is different from the other, there is no variation of return results form within the same website. So it just needs to analyze the definition of the DTD extraction in each e-commerce website, and save it into XML Schema, then utilize this form information to re-organize the search results of HTML pages' content into a unified form document, so that customers could browse the products information in a unified form, which enables customers to make comparisons with different products and select suitable products to make shopping decision. And it is important that, the definition of web page data crawling rules can be applied to other same type web pages with the sites [9].

In this study, we firstly defined a tree information structure for the web pages, then mapped the content of the pages to each corresponding field of the information structure (which is called as information attribute, and a part of the information structure). The process of describing information structure of web pages is just the process of analyzing what HTML document structure of the target web pages is. Assuming that the content of the target web pages is stored in database and the pages are generated dynamically, then this process is just as a reverse engineering of reconstruction of the database. The quantity of the elements and attributes of each HTML page may be very large frequently, so it needs several "analyzing - verifying - reanalyzing" cycle process to grasp the information structure of the target pages. This study utilized some structural marks in the HTML pages to locate the data to be captured. These marks could be the specific HTML marks or the attributes of HTML elements which are generally related to the specific display styles. As the specific display styles are always for the purpose of highlighting specific semantics, these marks are highly valuable. The rules of data capturing from specific website pages are formed through utilizing these marks to identify the starting and ending points of the specific data capturing.

Secondly, user preferences: To solve the issue how to filter information according to user preferences, this study uses multiple criteria decision making method of social choice theory. The user's preference is not static and constant, in fact, the preference of a person could be seen as a synthesis of criteria form which consists of more than one synthesis rule. The multiple criteria decision making methods can be used to solve the problem how to make a reasonable assessment of the products and then sort them, according to several conflicting criteria. Therefore, the general way to solve this problem is: After capturing the user's preference, the system will assess the products based on it, then sort the products and get a result list. When capturing the next batch of products information, the new products will get assessed with the products of former result list. According to the rules of multiple criteria decision making methods, it can be inferred that if the process of decision is not related to the independence of alternative objects, although it just assessed a subset of all of the products, the final result will not be effected.

The issue how to get user preferences is frequently discussed in recent years. In this study, we take this problem as a process of learning user preferences according to

user search records history. The system will record the customers' recent search history and shopping decisions to learn the customers' preferences. Thus, the stable preferences will be enhanced, while the preferences of rare tendency will be eliminated. By doing so, the system can keep close to customers' true orientation [10].

Thirdly, ranking products by key attributes: According to studies of customer decision-making factors in online shopping, the factors that customer s value most include: price, seller's reputation, buyers comments (word-of-mouth) and product description details [11]. So this study re-ranks the products in database in accordance with several attributes including price of products, seller's reputation and the number of positive reviews, then stores the ranking values. Meanwhile, the system also provides to the advanced users with the function that they could revise or reset the weight of each factor individually to make it closer to their true orientation. When customers browse products information, the system firstly displays the most suitable products list ranking by default rules based on the information in the user preferences database. Meanwhile the system also provides the options of ranking by price, seller's reputation and buyers comments, thus to realize the secondary information screening.

4 Conclusion

This paper researches Multi-Agent System for decision support of online shopping, of which specific target is to replace customer to search the huge information of commodities and sift useful information according with customer's purchase intention. In our work, we apply agent technology as the basis of system modeling. Taking the advantage of XML technology in information transfer, we achieve cross-site and automatically obtain network commodity information, which is a try of new e-commerce model and very convenient for user. However, this paper doesn't address the follow-up purchase and pay issues. Automatic payment forms the real e-business activities, of which the important issue is safety. How to make sure that all the business activities above are safe is a new problem we should seriously consider next.

References

1. China Internet Network Information Center (CNNIC). The 26th China Internet Development Statistics Report [EB/OL], <http://research.cnnic.cn/html/1279171593d2348.html>
2. Otterbacher, J.: Searching for Product Experience Attributes in Online Information Sources. In: ICIS 2008 Proceeding of International Conference on Information Systems (ICIS), Paris, pp. 207–210 (2008)
3. Nwana, H.S.: Software Agent: An Overview. Appear in Knowledge Engineering Review 11(3), 205–244 (1996)
4. Van Katwijk, R.T.: Multi-Agent Look-ahead Traffic-Adaptive Control. Technische Universiteit Delft, Netherlands (2008)
5. Cao, Y., Du, J., Bai, Y., et al.: Multi-layer Intelligent Encapsulation of Manufacturing Resources for MAS-based Distributed Manufacturing System. In: Second International Workshop on Knowledge Discovery and Data Mining (2009)

6. Zhang, Y.: Process Configuration Pattern: An Approach to the Modeling and Development of Multi-Agent Systems(MAS) Computer Software and Theory. Sun Yat-sen University (2004)
7. Ren, Z.: Design of AGPS Mobile Positioning System based on MAS. Fudan University (2007)
8. Pechoucek, M., Vladimir, M.V.: Industrial Deployment of Multi-Agent Technologies: Review and Selected Case Studies. *Autonomous Agents and Multi-Agent Systems* 17, 397–431 (2008)
9. Nayak, R., Iryadi, W.: XML schema clustering with semantic and hierarchical similarity measures. *Knowledge-Based System* 20(4), 336–349 (2007)
10. Kim, B.-D., Rossi, Peter, E.: Purchase Frequency, Sample Selection, and Price Sensitivity: The Heavy-User Bias. *Marketing Letters* 5(1), 57–67 (1994); p.11, 7 Charts
11. China iResearch Consulting Group. iResearch China Online Shopping Research Report (2009-2010) [EB/OL], <http://www.iresearch.com.cn/Report/1465.html>

Robot Real-Time Motion Planning and Collision Avoidance in Dynamically Changing Environments

Zhang Jin-xue

College of Electronic Engineering, Huaihai Institute of Technology, 222005, China

Abstract. Research in motion planning has been striving to develop faster and faster planning algorithms in order to be able to address a wider range of applications. The paper provides a literature review of previous works on robot motion planning and collision avoidance. A study on a novel force field method for robot motion planning has been given. Firstly, force field method is present, then, subgoal-guided force field method is proposed, finally, simulations on robot motion planning are carried out in the Player/Stage platform. The Subgoal-Guided Force Field method is suitable for real-time motion planning and collision avoidance in partially known and dynamically changing environments. Simulation results verify the feasibility and performance of the proposed methods.

Keywords: Motion Planning, Player/Stage, Collision Avoidance, Real-Time.

1 Introduction

Robot applications are increasingly being used in a variety of environments. Examples include home cleaning robots [1][2], automated wheelchairs to assist the handicapped or elderly people [4], museum-guide robots in museums and exhibitions [5][6][7], autonomous straddle carriers in container handling [8][9], driverless cars in the DARPA Grand Challenge [10][11], and so on. A basic requirement of fulfilling their tasks is that a robot must be able to move from its start point to destination and avoid possible collisions. This raises the problem of motion planning, which can be described as the construction of a collision-free trajectory that connects a robot to its destination [3]. Being a key challenge of robotics and automation engineering, motion planning has been studied extensively in past decades and a variety of approaches has been presented [3][13]. The robot motion planning and collision avoidance problem has been extensively studied in the past three decades and a variety of approaches has been developed [3].

Approaches to single robot motion planning may not be transferred directly to multi-robot motion planning and coordination. A team of robots is often utilized to accomplish complex tasks such as coordinated material handling [8][15], the exploration of unknown terrain or robot soccer [16][17].

Recent advances in the area of robot motion planning have resulted in the successful application of these techniques to diverse domains, such as assembly planning, virtual prototyping, drug design, and computer animation. Despite these advances, however, some areas of application have still remained out of reach for

automated planning algorithms. Applications requiring robots with many degrees of freedom to operate in highly dynamic and unpredictably changing environments fall into that category. In these domains a change in the environment can invalidate the previously planned path. To operate robustly and safely in dynamic environments the ability to modify the planned motion in real-time is necessary.

The problem of mobile robot motion planning and collaboration is addressed in this paper.

This paper is organized as follows. Section 2 gives the detailed related work done. Section 3 presents mobile robot motion model for our architecture. Section 4 gives the detailed description of robot motion planning method. Section 5 gives the simulation results and section 4 concludes the paper.

2 Related Work

The robot motion planning and collision avoidance problem has been extensively studied in the past three decades and a variety of approaches has been developed.

Motion planning is one of the fundamental problems in robotics. Broadly speaking, it is the problem of selecting a set of actions or controls which, upon execution, would take the robot from a given initial state to a given final state. Motion planning is a challenging problem whose complexity stems primarily from two sources: environment complexity and system complexity. The former refers to the complexity of planning trajectories for perhaps simple robots but in complex environments. Approaches developed for mobile robot motion planning may be broadly divided into three major categories: roadmap-based methods, cell decomposition-based methods and potential field-based methods [2].

Collision avoidance refers to the methodologies of shaping the robot's path to overcome unexpected obstacles. The resulting motion depends on the robot actual location and on the sensor readings. There are a rich variety of algorithms for obstacle avoidance from basic re-planning to reactive changes in the control strategy. Proposed techniques differ on the use of sensorial data and on the motion control strategies to overcome obstacles.

In this paper force field method is present and by combining the concept of subgoal with force field method, the subgoal-guided force field is proposed.

3 Mobile Robot Motion Model

In this section, a robot is modelled as a rigid body moving on a horizontal plane. The dimensionality of the robot in a working space is three, two for position in the plane and one for orientation along the perpendicular axis, which is orthogonal to the plane. Other factors which may affect the robot's movement, such as wheel slippages and additional degrees of freedom due to a robot's internal structure, are ignored.

Two reference frames are defined: a global reference frame $O(X, Y)$ and a local reference frame $o(x, y)$ attached to a robot's body, which is illustrated in Fig1. The axes X and Y define an inertial frame of reference on the plane as a global reference frame from an origin $O(X, Y)$. To specify the position of the robot, choose a point

o on the robot as its position reference point. $o(x, y)$ defines two axes relative to o on the robot body and is thus the robot's local reference frame. Note that x axis is set to be along this robot's moving direction. The position of o in the global reference frame is specified by (X_r, Y_r) , and the angular difference between the global and local reference frames is given by θ_r . The pose of a robot in the global reference frame is described as a vector with three elements.

$$r = [X_r \quad Y_r \quad \theta_r] \tag{1}$$

To describe the motion of a robot motion, it is necessary to map the robot motion along the axes of the robot's local reference frame to motion along the axes of the global reference frame. The mapping is a function of the current location of the robot and is accomplished using the orthogonal rotation matrix:

$$R(\theta_r) = \begin{bmatrix} \cos \theta_r & -\sin \theta_r \\ \sin \theta_r & \cos \theta_r \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} X_r \\ Y_r \end{bmatrix} \tag{2}$$

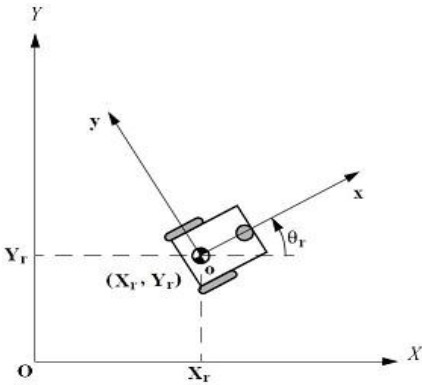


Fig. 1. Global reference frame and local reference frame

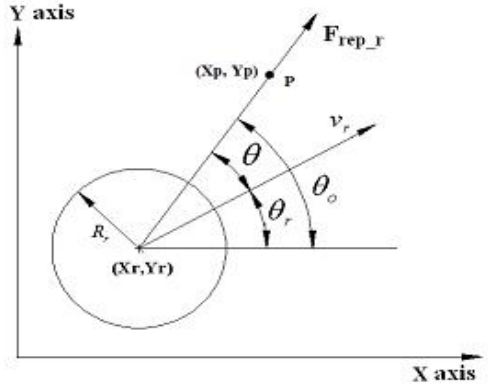


Fig. 2. Illustration of a robot's parameters

Consider a scenario of N mobile robots navigating in a global coordinate frame. Let robot i , at time $t = s$, locate at is (X_i, Y_i) and orient with angle θ_i to the X-axis of the world coordinate. The state of robot i can be described by:

$$r_{i,s} = [X_{i,s}, Y_{i,s}, \theta_{i,s}]^T, i = 1, 2, \dots, N \tag{3}$$

A robot is assigned a task such that it needs to travel from its current or start location to a goal location at $r_g = (X_g, Y_g)$. The robot then moves, under an appropriate control, towards the goal in a number of time steps Δt . The equation that describes the robot motion from time interval s to $s + 1$ can be expressed as a function of the current location, orientation and control:

$$r_{i,s+1} = h(r_{i,s}, u_{i,s}) \tag{4}$$

where $u_{i,s} = (v_{i,s}, \omega_{i,s})$ is the control issued to the mobile robot i , in which $v_{i,s}$ is the translational velocity and $\omega_{i,s}$ is its angular velocity. The motion equation of robot i at time interval $s + 1$ described by:

$$\begin{bmatrix} x_{i,s+1} \\ y_{i,s+1} \\ \theta_{i,s+1} \end{bmatrix} = \begin{bmatrix} x_{i,s} + v_{i,s} \cos(\theta_{i,s})\Delta t \\ y_{i,s} + v_{i,s} \sin(\theta_{i,s})\Delta t \\ \theta_{i,s} + \omega_{i,s}\Delta t \end{bmatrix} \tag{5}$$

The control commands, $v_{i,s}$ and $\omega_{i,s}$ are determined such that during the travel, the robot is free of collisions onto obstacles or other robots. The force field method is applied to generate the control commands based on the forces exerted on the robot, which is the function of attractive force from the goal and repulsive force from obstacles and other robots:

$$u = f(F_{rep_o}, F_{rep_j}, F_{att_g}) \tag{6}$$

where F_{rep_o} is the repulsive force from obstacles, F_{rep_j} is the repulsive force from other robots, and F_{att_g} is the attractive force to the goal. The control is further expressed by its components as

$$u = \begin{bmatrix} v \\ \omega \end{bmatrix} = \begin{bmatrix} f_v(F_{rep_o}, F_{rep_j}, F_{att_g}) \\ f_\omega(F_{rep_o}, F_{rep_j}, F_{att_g}) \end{bmatrix} \tag{7}$$

4 Robot Motion Planning Method

4.1 Definition of a Forcd Field

In the force field method, a force field is defined as a virtual field of repulsive force in the vicinity of a robot when it travels in a working space. The magnitude and orientation of a force field are determined by, and vary with, the robot’s status. This virtual repulsive force increases when the distance to the robot decreases. Note that the construction of the force field described below is in the global reference frame.

$$\theta = \theta_0 - \theta_r \tag{8}$$

$$E_r = \frac{v_r}{v_{\max} C} \tag{9}$$

$$D_{\max} = \frac{kE_r R_r}{1 - E_r \cos\theta} T^p \tag{10}$$

$$D_{\min} = \rho_0 D_{\max} \tag{11}$$

where θ_r denotes a robot's orientation in a global coordinate system. R_r is the radius, from the robot's origin, of the minimum circle embedding the robot entity. For any point $P(X_p, Y_p)$ in the robot's local reference frame, θ denotes the angle of this point in the robot's coordinates (Fig2), which can be determined from θ_r and the angle of this point in the global coordinates (θ_0). C is a positive number ($C > 1$) which represents the environmental influence to the force field. E_r is a positive decimal fraction with $0 \leq E_r < 1$. k is a positive coefficient which determines the size of the area to be covered by the force field. D_{max} is the maximum action distance of a robot's force field and D_{min} is the distance at which this robot has maximum repulsive force. D_{max} shows how far this robot can affect others in its vicinity. D_{min} provides a safe distance for the robot to prevent other objects from moving into this area. ρ_0 is a positive fractional number with $0 < \rho_0 < 1$ that defines how close the robot can be to obstacles. T_p represents the priority of a robot, which is especially useful for multi-robot coordination. Note that for single robot case, T_p is set to be 1.

The magnitude of a repulsive force to be generated by a robot's force field is defined by:

$$|F_{rep_r}| = \begin{cases} 0 & D \leq D_{min} \\ \frac{D_{max} - D}{D_{max} - D_{min}} F_{max} & D_{min} < D < D_{max} \end{cases} \tag{12}$$

where D is the shortest distance from point $P(X_p, Y_p)$ in the 2D space to the perimeter of the robot. P is a positive constant scalar which determines the magnitude of the repulsive force. When D changes from D_{min} to D_{max} , the magnitude of the repulsive force changes from P to 0 gradually. Furthermore, F_{max} is the maximum repulsive force which will cause the maximum deceleration on the robot.

4.2 Variable Speed Force Field Method

The Variable Speed Force Field method adopts the definitions of force field. In the Variable Speed Force Field method, a robot's motion is determined by the attractive force and repulsive forces acting on it. In particular, its translational acceleration is a function of the attractive force and repulsive forces acting on it and its angular acceleration is a function of force moments acting on it, as given by Newton's Second Law of Motion.

Define a fixed global reference frame $O(X, Y)$ and a moving local reference frame $o(x, y)$ attached to a robot's body. Let v be the robot's translational velocity and ω its angular velocity. Assume that a robot has mass m and inertia I about its centre of mass. Its translational acceleration \dot{v} and angular acceleration $\dot{\omega}$ in the moving frame $o(x, y)$ are given by:

$$m \dot{v} = F_{att_g} + \sum F_{rep_o} + \sum F_{rep_l} \quad (13)$$

$$I \dot{\omega} = M_{att_g} + \sum M_{rep_o} + \sum M_{rep_l} \quad (14)$$

where F_{att_g} is the attractive force from the goal position, F_{rep_o} are repulsive forces from obstacles and F_{rep_l} the repulsive forces from other robots. M_{att_g} , M_{rep_o} and M_{rep_l} are force moments generated by F_{att_g} , F_{rep_o} and F_{rep_l} , respectively.

4.3 Subgoal-Guided Force Field Method

The concept of subgoal is used in robot motion planning [12] [14]. In [12], the subgoal positions are continuously updated based on sensor data while robots are moving. In this section, a Subgoal-Guided Force Field method has been developed, in which the concept of subgoal is integrated with the Variable Speed Force Field method. The Subgoal-Guided Force Field method follows the definition of the force field and repulsive force of the Force Field method, but changes the way of defining attractive force. It is a generic approach for all Force Field based methods. In the scenarios considered, a robot is equipped with a laser sensor. This robot is capable of identifying openings in front of it based on the returned range data. The midpoints of these openings are selected as subgoal candidates. A heuristic function is then utilized to evaluate these candidates and the subgoal candidate with lowest cost will be selected as the current subgoal.

Fig3 illustrate how to determine subgoal candidates. Firstly, two openings are found by the laser sensor. Opening 1 is a gap between the front obstacle and the wall on the right and Opening 2 is a gap between the obstacle and the wall on the left. The midpoints of Opening 1 and 2 are denoted by P_1 and P_2 , respectively. P_1 and P_2 are selected as subgoal candidates.

The cost function applied by the robot to evaluate subgoals is defined as a sum of two functions. The first function is the distance from the robot position to a subgoal candidate (denoted by S_1). The second function is the estimated distance from a subgoal candidate to the global goal (denoted by S_2). In the scope of this research work, a robot is assumed to have knowledge of its own position in the globemap and the position of the subgoal candidate can be estimated from the sensor reading, so S_1 and S_2 can be calculated. The cost function is defined by:

$$g = k_1 s_1 + k_2 s_2 \quad (15)$$

The subgoal candidate with lowest cost g will be selected as the current subgoal. If there is more than one minima existing, the subgoal will be selected randomly from candidates with minimum cost. By tuning weighing factors $k_1 s_1$ and k_2 , different subgoals may be selected.

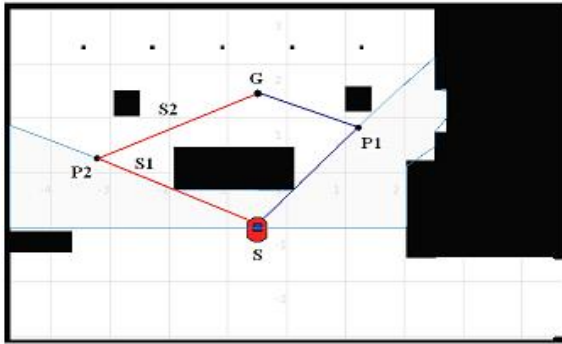


Fig. 3. Subgoal-Guided Field Method

5 Simulation Results

This section presents the simulations carried out using the Player/Stage platform [18]. A robot is supposed to travel from its start point to a goal point in indoor environments. This robot is equipped with a laser sensor and an Adaptive Monte-Carlo Localization driver (amcl) [19] is utilized to determine its location. The update threshold of the amcl driver is set to be 10 degrees or 5cm, that is, the amcl driver returns with this robot's new pose when there is a displacement of 5 cm or angular displacement compared to the amcl driver's last output. Robot parameters are selected based on those of a pioneer robot [20]. In all simulations, k_1 and k_2 are set to 1.

5.1 Experiment Setup

5.1.1 Software Platform

The Player Project (formerly the Player/Stage Project or Player/Stage/Gazebo Project), which is a project for the testing and development of robotics and sensor systems, has been adopted to develop software in experimental studies [18]. Its components include the Player network server and Stage and Gazebo robot platform simulators. The project was developed in 2000 by Brian Gerkey, Richard Vaughan and Andrew Howard and is widely used in robotics research and education.

A. Player Robot Device Interface

The Player is a TCP/IP based network server for robot control. It supports a variety of robot hardware. Player is designed to be language and platform independent, which means that the client program can be written in any language that supports TCP sockets, like C++, Java and Python, etc. In the simulations and experiments to be presented in this chapter, the control softwares are written in C++. Player is also designed to support any number of clients, which makes it suitable for complex applications such as multi-robot motion planning and coordination.

B. Stage Multiple Robot Simulator

The Stage has been designed to be capable of simulating a population of mobile robots moving in a two-dimensional bitmapped environment. Various sensor models

are provided in this platform, including sonar, scanning laser rangefinder, pan-tilt-zoom camera with colour blob detection and odometry. Stage devices present a standard Player interface so that very few or no changes are required to move between simulation and hardware. That means that if control software has been tested successfully in the Player/Stage, it is almost ready to work on real robots [18].

5.2 Simulation

The map used in this simulation is shown in Fig4. A robot is supposed to travel from the start point at $S(-3, -2)$ to the destination $G(1, 1.5)$. Fig 4 gives a snapshot when the robot is about to move. Six subgoal candidates are found (denoted by $P_1, P_2, P_3, P_4, P_5, P_6$) here. P_2 is selected as current subgoal. The resultant path is shown in Fig 5.

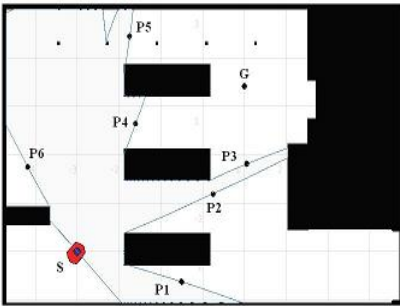


Fig. 4. Subgoal-Guided Field

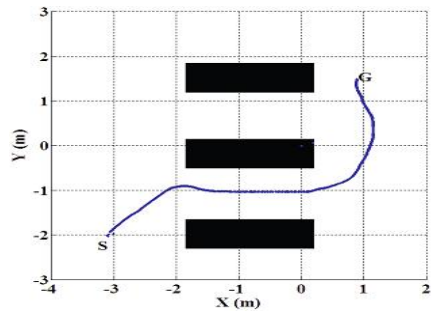


Fig. 5. Subgoal-Guided Field Method – map Method- resultant path

If the environment changes, for example, the lower corridor is blocked as shown in Fig6, the subgoal candidates found are depicted as P_1 to P_4 . These simulations show that the Subgoal-Guided Force Field method is applicable in dynamically changing environments.

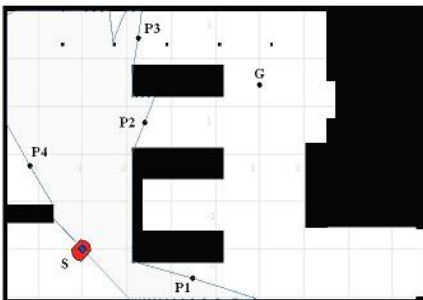


Fig. 6. Subgoal-Guided Field Method

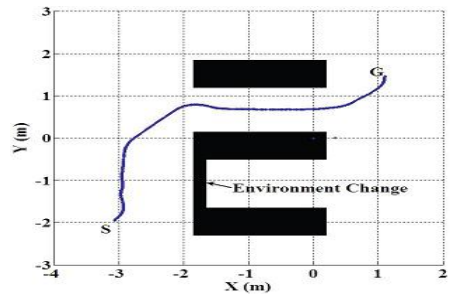


Fig. 7. Subgoal-Guided Field Method - new path - environment changed

Simulations show The Subgoal-Guided Force Field method is suitable for real-time motion planning and collision avoidance in partially known and dynamically changing environments.

6 Conclusions Results

This paper has presented a Subgoal-Guided Force Field method, which greatly improves the performance of the Force Field method. In the Subgoal-Guided Force Field method method, a robot first analyses the surrounding environment and finds possible passages to approach. The proposed Subgoal-Guided Force Field method has been demonstrated feasible and efficient by simulations carried out in the Player/Stage platform.

In the future work, we will extend the proposed method for various types of robots and extend it from 2D to 3D. In addition, the research will include not consider feedback control, which will be researched in the future.

References

1. Juang, J.G.: Robot collision avoidance control using distance computation. In: Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, vol. 3, pp. 2564–2569 (1995)
2. Jones, J.L.: Robots at the tipping point: the road to iRobot Roomba. *IEEE Robotics & Automation Magazine* 13, 76–78 (2006)
3. Latombe, J.-C.: Robot motion planning, Kluwer Academic Publishers, Boston (1991)
4. Taha, T., Miro, J.V., Liu, D.: An efficient path planner for large mobile platforms in cluttered environments. In: Proceedings of the IEEE International Conference on Robotics, Automation and Mechatronics, Bangkok, Thailand, pp. 225–230 (2006)
5. Nourbakhsh, I.R., Kunz, C., Willeke, T.: The mobot museum robot installations: a five year experiment. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3636–3641 (2003)
6. Thrun, S., Bennewitz, M., Burgard, W., Cremers, A.B., Dellaert, F., Fox, D., Hahnel, D., Rosenberg, C., Roy, N., Schulte, J., Schulz, D.: MINERVA: a second-generation museum tour-guide robot. In: Proceedings of the IEEE International Conference on Robotics and Automation, pp. 1999–2005 (1999)
7. Bennewitz, M.: Mobile robot navigation in dynamic environments, PhD Thesis, University of Freiburg (2004)
8. Durrant-Whyte, H., Pagac, D., Rogers, B., Stevens, M., Nelmes, G.: Field and service applications - an autonomous straddle carrier for movement of shipping containers - from research to operational autonomous Systems. *IEEE Robotics & Automation Magazine* 14, 14–23 (2007)
9. Patrick: technology & systems: autostrad, <http://www.patrick.com.au/IRM/Content/technology/autostrad.html> (accessed December 20, 2008)
10. Hoffmann, G.M., Tomlin, C.J., Montemerlo, M., Thrun, S.: Autonomous automobile trajectory tracking for off-road driving: controller design, experimental validation and racing. In: Proceedings of the 2007 American Control Conference, pp. 2296–2301 (2007)

11. DARPA Grand Challenge Information, <http://www.darpa.org/> (accessed at December 2008)
12. Chen, P.C., Hwang, Y.K.: SANDROS: a dynamic graph search algorithm for motion planning. *IEEE Transactions on Robotics and Automation* 14, 390–403 (1998)
13. Hwang, Y.K., Ahuja, N.: Gross motion planning - a survey. *ACM Computing Surveys* 24 (1992)
14. Zhu, D.J., Latombe, J.C.: New heuristic algorithms for efficient hierarchical path planning. *IEEE Transactions on Robotics and Automation* 7, 9–20 (1991)
15. Liu, D.K., Kulatunga, A.K.: Simultaneous planning and scheduling for multi autonomous vehicles. In: Dahal, K.P., Tan, K.C., Cowling, P.I. (eds.) *Evolutionary Scheduling*, pp. 437–464. Springer, Heidelberg (2007)
16. RoboCup Official Site, <http://www.robocup.org/> (accessed December 20, 2008)
17. Kitano, H., Suzuki, S., Akita, J.: RoboCup Jr.: RoboCup for edutainment. In: *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 807–812 (2000)
18. Gerkey, B.P., Vaughan, R.T., Howard, A.: The Player/Stage project: tools for multi-robot and distributed sensor systems. In: *Proceedings of the International Conference on Advanced Robotics, Coimbra, Portugal*, pp. 317–323 (2003)
19. Fox, D.: Adapting the sample size in particle filters through KLD-sampling. *The International Journal of Robotics Research* 22 (2003)
20. Pioneer robot, <http://www.activrobots.com/ROBOTS/p2dx.html> (accessed December 20, 2008)

Design of the On-Line Detection System for the Surface of the Hot Heavy Rail

Renxi Gong and Yuxia Liu

School of Electrical Engineering, Guangxi University,
530004 Nanning, China
rxgong@gxu.edu.cn, liuyuxia230@163.com

Abstract. A construction of an on-line detection system for surface of the hot heavy rail is presented, and a kind of detection means based on the machine vision technology is proposed. In the paper, a description of the principle of machine vision detection, the operational principle of CCD and the construction of the system is made. The overall design of the system, including light controls, image acquisition, image processing and man-machine interface management is described in a modularized form. The system has been practically carried out in a Rail Beam Plant. The client in the local area network could observe the image, make information analysis by on-line or off-line, and manage lots of information. Good results have been obtained.

Keywords: Hot heavy rail, surface quality, machine vision, image acquisition, online detection.

1 Introduction

Heavy rails have been one of the key products of steel enterprises all long. They have brought huge economic benefits for the enterprises. The quality of the heavy rails is not only an important factor of ensuring the safe transportation of railway, but also the key factors of market competition. The quality of the rails directly influences the safety of the national properties and people's lives. With the ultrafast development of the high-speed railway projects and the heavy load of the rail, the better and better quality, function and performance of the heavy rail is required. Among them, the surface quality of the heavy rail is one of the extremely key points.

At present, the main method about the detection of the surface of heavy rail in the rail beam plant of several enterprises is to eyeballing method in which the quality of the rail is judged by the observation of the workers. A blaze of light source is usually used to shine on the surface of the heavy rail and the reflection of the surface is observed by the workers to determine whether the defects exist or not. When the heavy rail is being rolled, it is in a state of high temperature and high radiation, and moves at the high speed. Owing to this bad condition of detection, many doubtful defects can't be given to draw conclusion in time. Moreover, the heavy rail has a few surfaces not be detected only by means of experience. Even if the detection is supplemented later on, the undetected regions absolutely exist. As a result, waste

products turn up frequently and it brings pecuniary lose that can't be avoided to enterprises.

Nowadays, with the great progress of science and technology, computer technology gradually comes to maturity. In order to obtain the information of the surface defects of heavy rail at the real time, an idea based on machine vision is brought up. The method of direct observation by naked eyes is thrown away. An on-line detection system based on the detection principle of machine vision is designed. Below is the detail description of the construction and the operation in the system.

2 The Construction and the Principle of the System

2.1 The Principle of Machine Vision

Machine vision system is a vision detection system which is based on machine vision technology and is mainly used to imitate people's vision by computers. It has been widely used in several fields such as production, manufacture and so on. In the system, the information is extracted from the images of objective things, then disposed, perceived and used for actual detection, measurement and control finally. In a simple word, it makes use of machines instead of people's eyes to do various measurements and judgement. The superiority of using this technology is that a very wide scope of optical spectrum response, large amount of information can be obtained and the constant, stable, reliable and repeated detection can be carried out in the comparatively bad condition.

2.2 The Principle of CCD

The primary detection device of machine vision system is CCD which is a solid camera device. Different from other majority of devices in which electric current or voltage is used as their operation signals, the electric charges act as its operation signal in CCD devices. Their fundamental functions are to produce, store, transmit and detect signal charges. While the lights shine on the photosensitive surfaces of CCDs, the transformation from light to electricity is firstly completed, which means that the electric charges that are proportional to the radiations of light will be generated. The objects that are radiated will reflect lights to CCD devices. Charges are gathered based on the bright and dark lights and then CCD produces corresponding feeble voltage signals which is directly proportional to the amount of light electric charges. After filtering and amplification, an electric signal or standard video frequency signal which can express intensity of the lights is output through driving circuit.

2.3 The Principle of the System

Generally speaking, the main construction parts of machine vision system includes cameral lens, light sources, industrial cameras and industrial computers. The constitution principle of this system is to stall four image collecting units behind the production line of heavy rail which can obtain the images of the surfaces of the top waist, bottom waist, bottom and tread of heavy rail. The simple illustration on the constitution of the system is shown in Figure 1.

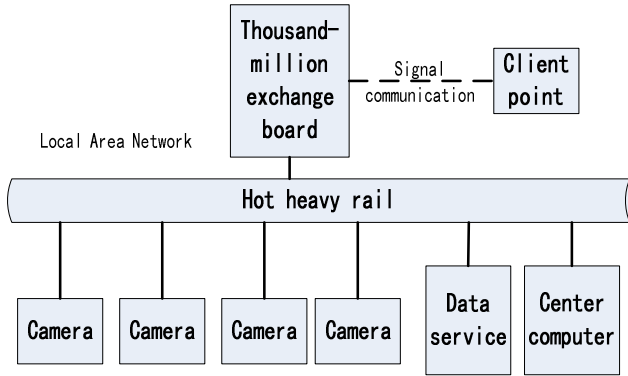


Fig. 1. The simple illustration on the constitution of the system

The running process of the system is as follows: the detector will send trigger impulses to the image collecting units when it finds out the objects approaching to the center of the visual field of the shooting system. The cameras are in a waiting state before the starting impulses come, and they begin to scan after the starting impulses come. The cameras open their exposed mechanisms before a new circulation of scan. The exposed time can be set beforehand. The cameras begin to scan formally and output the images after exposing. It means that the acquisition of images about heavy rail surfaces begins. Then the collected imitative image data will be transformed into digital image data. The digital data will be saved in data base after an operation by computers. The whole system composes a local area network. The final data are sent into central computers through the thousand-million exchange board in this local area network for detecting. After an operation, the data are sent into the data service. At the same time, the detected images will be shown on big screens of the central computer so that the workers can make a judgement. The states at the real time of the surface quality of heavy rail can be observed in the clients of the local area network.

3 The Modularized Design of the System

3.1 The Part of Light Sources

The shooting quality of cameras mainly relies on the assistant light sources. The primary principle is that the lights from light sources shine on the chief shooting surfaces and to put quite many reflecting lights to the full into the cameras. Due to the extremely high temperature of the hot heavy rail and the extremely low reflecting rate of the surfaces of hot heavy rail themselves, the intensity of light sources are demanded to be higher. Beside of the requirement of live high-speed shooting, the selected light source has to satisfy the requirement above. Any frequency flashes of light sources themselves are not allowed. It means that the lights must shine steadily.

According to the requirements above, the types of light sources should be selected. The chief characteristics of light sources are looked up and shown in Table 1 below.

Table 1. The characteristics of commonly used light sources

Name of Lamp	The consumption electricity (W)	Voltage (V)	Using time(h)
Tungsten	15~200	200	3000
Halogen	100	220	3000
Neon	16	220	6000
Daylight	4~100	220	500~8000
LED	extremely low	very low	1000

After analysis, it can be found that only neon lamp could satisfy the requirements of the intensity, stability and high-speed shooting. Therefore, the neon lamps are used in the whole system. This kind of light source can illuminate steadily. Its intensity conforms to the requirement of the shooting intensity of the camera. And its setting satisfies the requests of the high temperature at the scene. It can illuminate steadily under the condition of high radiation of the hot heavy rail, thus effectively decreasing the changing period of the lamps and meeting the realistic need at the real place.

3.2 The Part of Image Acquisition and Processing

The heavy rails themselves may radiate infrared rays when they are in the high temperature state. The brightness of images is easily saturated and the images are not clear even if the stop of the camera is small. Moreover, the radiation of a few invisible lights may also influence the image quality. Every means must be tried to get rid of the disturbing lights. The function of the filter is to filter out the lights that are meaningless to the formation of image. As a result, in this system, the filters are added on the camera lens at first which are used to filter out those disturbing lights to improve the image quality.

The imaging quality is affected not only by the shining angle of assistant light sources, but also by the parameters of the camera lens and cameras. The parameters of the camera lens include the distance of the camera from the objects, and the value of the stop. The parameters of the camera include the white balance parameters, system gain, shooting frequency and so on. The setting of the parameters of the camera lens is used for ensuring the surfaces of heavy rail to be in the range of the depth of field. By adjusting the parameters of the camera, the setting of the colors, brightness and shooting frequency of the images output by the camera can be made. According to the detecting need at the scene, a type of colorful line-scan CCD camera is used in the system.

The surfaces of the heavy rail are complex with many planes and curved surfaces. A single camera obviously can't finish the image acquisition of all the surfaces. The heavy rail moves at a certain speed. Taking account of the undulation of the curve surfaces of heavy rails when they are moving, for the sake of on-line detecting of the surfaces of heavy rails, in accordance with the shape characteristics of the surfaces of

heavy rails and the former detecting experience of the enterprises, four separated collecting units which are composed of line-scan CCD cameras, camera lens and protecting covers with blowing installations are used in this system. Images are collected from the four parts of the heavy rail which usually appear defects, the surfaces of the top waist, bottom waist, bottom and tread.

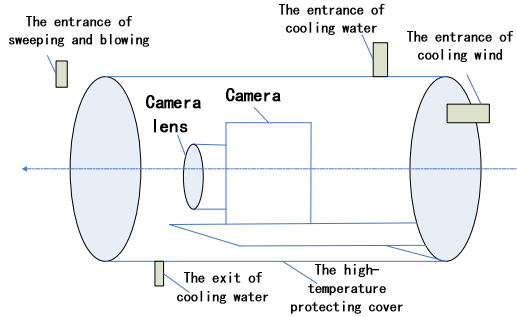


Fig. 2. The separated collecting unit

The simple diagram of the separated collecting unit is shown in Figure 2. The camera is fixed in the protecting cover. The protecting cover links the water pipes and the windpipe. It is used for ensuring the temperature to be at the given value when the camera is running in order for it to operate steadily and persistently. The windpipe joins to the side of the front part of the protecting cover which is used for sweeping the dust on the heat insulating glass and diminishing the influences of the image quality by the dust. All the angles of holders fixed for the camera in each separated collecting unit can be adjusted freely so that it is ensured the distance and the shooting angle can change in a rather large range.

The production line of heavy rail is on line in the 24 hours of the whole day. And it is an on-line detecting system. Therefore, a great number of image data need to be saved in the computer and the saving must be in a high-speed way. For the transmission of image signals need to be in a quite high speed, the commonly used transmission interface can't satisfy the requirements. In order to obtain the defect information of the surfaces, the image capture cards are needed. The main function of the image capture card is to capture the video frequency data output by the camera in time and offer the high-speed interfaces with computers. The image capture card of the Matrox Company is used in this system. It has four thousand-million network interfaces which can link with cameras at the same time. It can control and detect the whole process of image acquisition effectively and reliably.

The heavy rail moves at a high speed. The formal velocity is ordinarily at about 4 to 5 meters per a second and not steady. In this case, the linear velocity of the surface is different when the heavy rail is being rolled. It means that the variation of velocity is closely related to the materials of heavy rail and the rolling standards in the process of rolling. It shows that the linear velocity is not steady at the certain collecting point. Therefore, the line-scan CCD is used for image sampling which could adapt to different velocity of the rolling heavy rail.

For the unit, it will greatly influence the formation of image if one can't get the exact linear velocity in time. After the image acquisition of the surfaces of heavy rail, for the sake of observation and analysis, the images need to be processed. The image process includes the image gray, image enhancement, image denoising, threshold segmentation, edge detection and so on. The collected colorful images should be firstly transformed into gray ones. Then the grayscale images are processed to get rid of the interference of noise by the method of image smoothing and sharpening. Finally, the methods of image segmentation and recognition are utilized. In this way, the clear images of the surfaces of the heavy rail can be obtained after the collected image data are processed.

3.3 The Part of the Human-Computer Interface Administration

It's for on-line detection of the surface quality of the heavy rail in this system. The collecting images are firstly distinguished and analyzed by the related technical workers with their eyes. And then they make judgements on whether the surface quality of the heavy rail can accord with the related producing requests. The types of the defects can be judged finally.

Based on the programming technology facing objects, the programming of the software system is built on the platform of Microsoft Visual C++6.0 which can achieve the effective administration of the image data, the information of the system and the devices. The image data which have no defects are acquiescently saved for 10 days in the system. And the data will be automatically overwritten with new ones. But the image data which are marked to be defective will be saved permanently in the system. Therefore, the human-computer interface administration is a important function in this system. By means of the research and improvement of the administration, the functions such as on-line image detecting of the hot heavy rail, defect marking and past image broadcasting are achieved. The operation is easier than before and the detecting efficiency is improved. It has achieved good results finally.

4 Results

The system has been practically carried out on the production line of the heavy rail in a Rail Beam Plant of a steel group. The image data of all the four surfaces which are obtained at the scene in time will be saved in the system. All the figures below are the cut images that are obtained at at the scene. In Fig.3 and Fig.4 is separately shown the clear images of the four surfaces of the heavy rail which are not defecting ones. Among them, the left image in figure 4 is the surface image of the top waist, the right image in figure 3 is the surface image of the bottom waist; the left image in figure 4 is the surface image of the tread, the right image in figure 4 is the surface image of the bottom. The comparison of images between the surfaces of the top waist is shown in figure 5. It can be seen that the there're obvious gaps on the surface in the right image and after they found the defects, the related technical workers at the real place stopped the production of the batch of the heavy rails in time. And then the waste ratio of the products decreases. It can be seen the defects are effectively detected at the real place by the system.

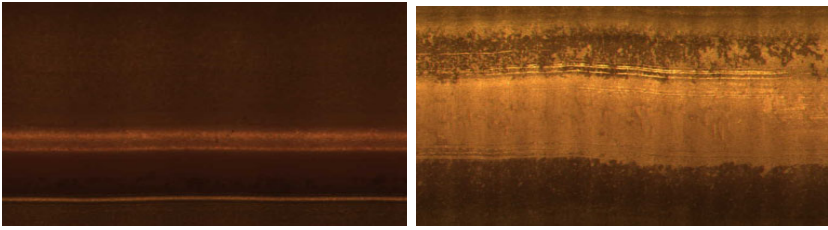


Fig. 3. The surfaces of the top waist and the bottom waist

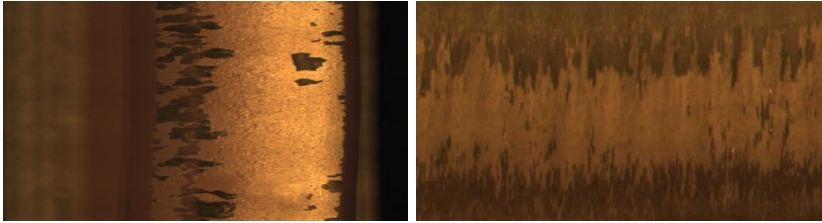


Fig. 4. The surfaces of the bottom and the tread

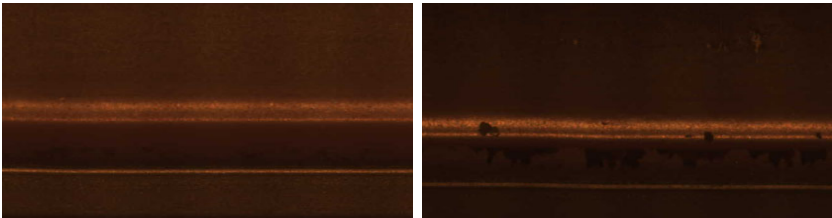


Fig. 5. The comparison of images between the surfaces of the top waist

5 Conclusion

An on-line detection system for surface quality of the hot heavy rail is designed in this paper. The clear images of the heavy rail are gained in the system by the methods of image acquisition and processing. The information of the surfaces of the heavy rail can be detected at the real time and the surface quality can be judged by the related technical workers through the human-computer interface administration. A good deal detecting abuses by workers are conquered and on-line detecting is accomplished in the system. The workload of detecting is lightened and it is good for finding the unqualified products as soon as possible. Thus, corresponding measures can be carried out and the follow-up process of the waste products which is so expensive can be reduced. At the same time, the producing efficiency and the quality of the heavy rail are greatly improved.

References

1. Xu, G.: The Application of Curve Fitting Technique in Fabric Defect Detection. In: 2010 IEEE International Conference on Signal Processing System (2010)
2. Hu, L., Duan, F.: Study on In-line Surface Detect Detection System for Steel Strip Based on Linear CCD Sensor. *Acta Metrologia Sinica* 26 (2005)
3. Li, J., Deng, W., Liu, L.: Adjustment and Calibration Method for Linear CCD Measurement System. *Modern Electronics Technique* 32 (2009)
4. Hu, L., Duan, F.: The Application of FPGA-Based Image Processing Technology in Surface Defect Detection System for Steel Strip. *Chinese Journal of Sensors and Actuators* 19 (2006)
5. Xia, C., Li, C., Xiong, C.-y.: Summary of The Online Test Methods of Surface Finish for Rolled Steel Plates. *Industrial Measurement* 20 (2010)
6. Hu, L., Duan, F., Ding, K.: Research on Surface Detects on Line Detection System for Steel Plate Using Computer Vision. *Iron & Steel* 40 (2005)
7. Li, Y., Wang, B.: Research on online measuring system of ceramic tile size based on linear array CCD. *Transducer and Microsystem Technologies* 29 (2010)

Application Research of Predictive Control in Heating System

Jinhong Wei, Heqiang Yuan, and Qianqian Wang

Institute of Mechatronics Engineering, Taiyuan University of Technology,
Taiyuan 030024, China
weijinhong@tyut.edu.cn, da4301@126.com, wqq36@139.com

Abstract. Though the traditional gas boiler system is both energy saving and environment protecting, there exists certain problems have to be solved. Therefore, A new method for settling the unreasonable control system is presented. The improved control scheme based on DMC-PID algorithm is proposed by analyzing the characteristics of the gas heating system, establishing the boiler system model and studying the characteristics of heating control system. Combined with the actual heating process, the algorithm and the implementation process are presented in detail. Finally, the performance of the algorithm is verified by use of simulation test. The results turn out that the improved algorithm has better efficiency than that of traditional algorithm, and thus has higher utility value.

Keywords: DMC Control, Matlab, DMC-PID, Hardware in-the-loop Simulation.

1 Introduction

With the continuous development of our economy and society, the demand of the people for the environment has increased significantly. The structure of energy has changed greatly. For the excellent environmental performance, commercial gas heating boiler has higher social benefits. With the adjustment of urban energy and the increased demand for the environmental protection, the traditional coal burning boiler has been gradually replaced by gas-fired boiler [1].

However, the existing gas heating systems among which there have significant differences in energy consumption and control effect have many shortcomings. Therefore, the usage of reasonable and advanced control methods is essential. For the large time delay property of secondary network of gas boiler heating system, the conventional PID control algorithm has certain limitations. With regard to nonlinear secondary network heating system, the PID cannot achieve better control effect due to its difficulties in parameters adjustment and lower control performance. The predictive control scheme is applied to large time delay and complex system, such as the nonlinear secondary network heating system, and could overcome the uncertainty of the system[2-3].

2 System Overview

The group control system of gas hot water boiler is divided into three parts: heat supply source, heat supply network and terminal users. Heat source produce heat medium, heat network transmit heat medium and the terminal user refers to the heat-usage site. Terminal user of the heating system includes heating system, hot water supply system and so on. Figure 1 shows the schematic diagram of gas boiler heating system. The system consists of nine 1T gas boiler, high district heating, low district heating, low and high areas hot water, water segregator and water collector. In secondary network control system of gas heating boiler, the intensity of the secondary heating network is enhanced or reduced by controlling the three-way valve, and thus the design requirements of the users' room temperature is achieved.

Thus, the control of the opening of three-way valve is an important stage. In general, the room temperature is controlled uses deviation between indoor real-time temperature and setting point by which the amount of heat is determined. Therefore, the control is implemented just after the system behavior has occurred, and is not timely. According to the basic idea of predictive control, the system behavior should be forecasted firstly, and the predictive value and expectations are compared. Then on the basis of the proper bias, the system could be controlled in advance by use of appropriate control scheme. Thus, the control method is intrinsically pre-control due to which the excess of settled range of the room temperature could be avoided. For gas boiler heating system, which is a complex structure, large time delay and nonlinear time-varying system, predictive control can meet the requirements of real time, adaptability and accuracy.

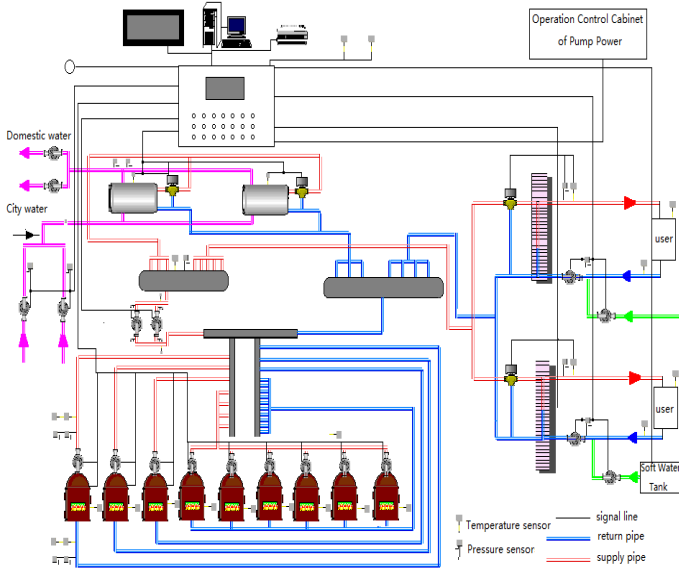


Fig. 1. Schematic Diagram of Gas Boiler Heating System

The control method is chosen according to actual situation on the basis of which the load characteristic of the system is analyzed and the control algorithm is determined[4].

3 The Study of Heating Control System Algorithm

3.1 Analysis of Heating Process Model

The heating process of gas boiler is the process of delivering heat generated by burning natural gas to the terminal users.

The transfer functions of the primary network and heat exchange:

$$t_{si}(s) = \frac{K_n}{T_n s + 1} Q_1 + \frac{1}{T_s s + 1} t_{so} \tag{1}$$

The transfer functions of heat dissipation of the secondary network:

$$t_{so}(s) = \frac{K_m}{T_m s + (K_m + 1)} t_{si}(s) + t_{out}(s) \tag{2}$$

Then, combined with the two equations above, the system transfer function can be obtained, as shown in (3).

$$t_{so} = \frac{K_n K_m}{T_n T_m s^2 + [(K_m + 1)T_n + T_m]s + 1} Q_1(s) + \frac{K_m(T_n s + 1)}{\{T_n T_m s^2 + [(K_m + 1)T_n + T_m]s + 1\} [T_m s + (K_m + 1)]} t_{out}(s) \tag{3}$$

Where: t_{si} -supply water temperature of secondary network, t_{so} -return water temperature of secondary network, Q_1 - water flow in primary network, t_{out} - outdoor temperature, k_n - the integrated scale factor of primary network, T_n - the integrated time factor of primary network, k_m - the integrated scale factor of secondary network, T_m - the integration time factor of secondary network. The system is a second-order system, and the interference is a third-order system with a first-order zero. Taking t_{so} as controlled variable and Q_1 as control variable. In the system, the most important is to control the opening of electric three-way valve in the heat transfer station, thus the size of hot water supply can be changed, and then the return water temperature of secondary network can be controlled in a relatively stable range.

3.2 Research on Algorithm

PID controller algorithm is simple, easy to adjust parameters and has the advantages of strong anti-interference, while DMC can deal directly with the controlled object which has the property of time delay, and has better tracking performance and also has strong robustness characteristics on the model mismatch[5-6]. Therefore, for the control of electric three-way valve, the inner loop uses conventional PID controller. After reasonable setting, the model can be combined to constitute the controlled object of the outside loop. DMC controller can be used in the outer loop to overcome

large time delay, large inertia and uncertainty of the object model. Outer loop takes the forecasted return water temperature as the main feedback signal to predict the changing direction and amount of return water temperature and to achieve adjusting the primary network flow ahead of regulation.

DMC-PID control structure of the heating system is shown in Figure 2. Outer loop uses dynamic matrix control, while inner loop uses PID control. $r(k)$ is the return water temperature reference trajectory, $u(k)$ is the control output, $G(z)$ is the heating system model, K_a is three-way valve opening scale factor, W is the influence of disturbances, K_v and K_H are the feedback loop of the opening and the return water temperature.

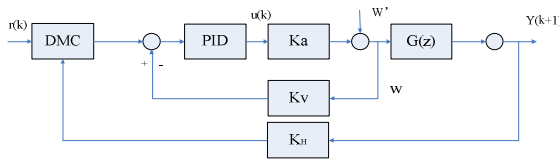


Fig. 2. DMC-PID control structure of boiler system

Taking Gauss low-pass filter for example, its two-dimensional form is shown as below equation.

DMC controller specific implementation is as follows:

(1) To implement dynamic matrix control in boiler heating system, in the first step, a single step response signal should be measured, and the step response model should be established. Actually, under normal operation, by using online identification method, the step response model of the controlled object is acquired by directly inputting the step signal to it. The water flow in primary network is Control variable, while return water temperature of secondary network is controlled variable. The device controlling primary network water flow is the electric three-way valve, return water temperature is acquired by temperature sensor. Therefore, the valve opening is determined to be the step response signal. The sampling value is the secondary network return water temperature $t_i = t(iT)$, where $i = 1, 2, \dots, T$ is the sampling period.

According to the analysis, the system is a second-order system. On the basis of the step response model $t_i = t(iT)$, the characteristics of the system can be analyzed by using MATLAB's system analysis tools.

(2) Assume that the current time is t , the control has to predict the electric valve opening increment $\Delta u(k), \Delta u(k + 1), \dots, \Delta u(k + M)$ of the future M moments from the current under the role of which the system output i.e. the future output prediction value $T_m T_m = (k + i | k)$ of the secondary network return water temperature T should

be as close as possible to the given expectation $a(k+i), i=1,2,\dots,n$. The optimization process is shown in Figure 3.

(3)Because of the presence of interference and the model disorder, the predicted value is likely not the same with the actual value and there will be a deviation. DMC control algorithm applies the feedback correction method to solve this problem, the actual temperature $T(k+1)$ is sampled and tested at every moment, and compare it with the prediction temperature output and get the deviation.

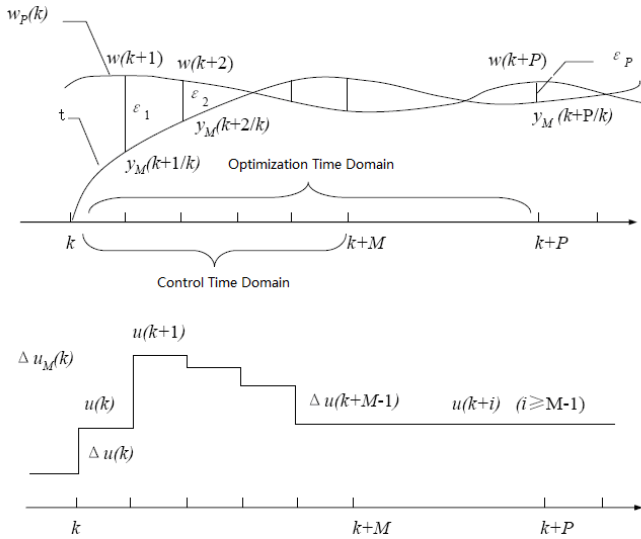


Fig. 3. Optimization process

(4)The only usage of instant deviation $T_e(k)$ to correct output prediction P could result in leakage of information. The model deficiencies can be improved by using the P-step prediction deviation. First, by differencing model error sequence to get $\omega(k) = T_e(k) - T_e(k-1)$ which then can be transformed into stationary random sequence. Random sequence $\omega(k)$ can be described by ARMA (n, m) model, as shown in (4).

$$\begin{aligned} &\omega(k) + a_1\omega(k) + a_1\omega(k) + \dots + a_1\omega(k) \\ &= \varepsilon(k) + b_1\varepsilon(k) + b_2\varepsilon(k) + \dots + b_p\varepsilon(k) \end{aligned} \quad (4)$$

Where $\omega(k)$ is a white noise sequence which has zero mean value and a variance δ^2 . Unknown parameters can be estimated by using recursive least squares method. Then, the forecast error can be obtained by adding together optimal predictive value and the above error, as shown in (5).

$$T'_{e(k+ik)} = \sum_{j=1}^i \varepsilon(k+i|k) + T_e(k) \quad (5)$$

4 Implementation of the System

The DMC-PID control is achieved in the controller in the host and lower computer respectively. The achievement of PID control in the controller needs parameter-settlement which will not be discussed here.

Based on the method of compiling dynamic link library (DLL), the mixed programming is achieved in which VB calls the DLL generated by MATLAB. Thus, the programmed VB application needs not the running of MATLAB, so the algorithm and VB program are integrated to become one.

The simple implementation process is: first, set comtool and choose an external compiler. Open a control file created by MATLAB and run comtool. Construct a new project named DMC, and the class name is DMCCCLASS. Then set the path and generate projects. Put M file into the path folder after these preparations. And then click add file in the comtool, add the saved. M file above, click build - com object, and thus the dynamic link library is generated.

Accordingly, in the course of VB programming, set the program name according to the related content of Matlab dynamic link library, and the return value is a reading value of Matlab dynamic link library, and reading value is the return value of Matlab dynamic link library.

5 Hardware-in-the-Loop Simulation

Before the PCD controller has not been installed in the actual control condition, the performance of heating system with PCD controller can be verified through hardware-in-the-loop simulation using computer interface card system. By using the principle of hardware-in-the-loop simulation[7], the hardware-in-the-loop simulation of boiler control system is achieved as shown in Figure 4. Practical application equipment of the control system is SAIA PCD2. Through the actual communication between PCD controllers and computer interface card system, the operation of the system is simulated. If the PCD control system performance does not meet design requirements, adjust the controller parameters or programs. Meanwhile, It is of great significance in setting the controller, the program modification and program test etc. The ultimate goal is to ensure that the controller and simulation systems can run under the closed-loop condition, to verify that the controller can achieve the effect of automatic control which determines how to modify the control program to achieve the best control effect.

The analysis of hardware-in-the-loop simulation experimental loop system comes as follows. Firstly, the hardware-in-the-loop simulation platform is designed according to the number of points required in the hardware-in-the-loop simulation. In the high-District second network experiment, there have to set an analog input of a three-way valve opening and five analog outputs which are supply water temperature

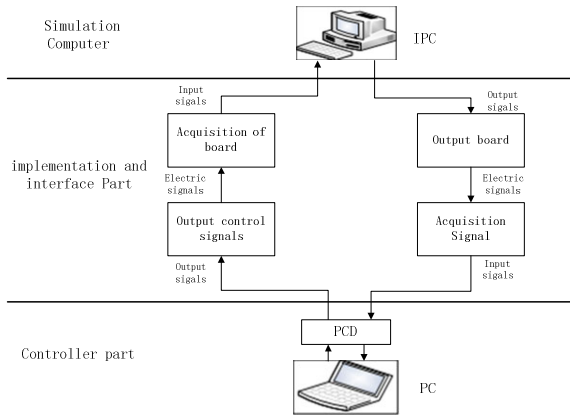


Fig. 4. Schematic diagram of the hardware-in-the-loop simulation design

of secondary network, return water temperature of the secondary network, indoor temperature, outdoor temperature, and hot water tank temperature.

Advantech PCL-818L data acquisition card provides 8 differential analog inputs, thus could meet the requirement of analog inputs. PCL-836 can use 3 PWM outputs as the analog signal source. PCL-728 provides 2 12-bit analog outputs. The simulation computer use Advantech industrial computer to connect the data acquisition card and create system model. The controller of physical part is the SAIA's PCD2 controller; the configuration program of controller monitors installs on PC side, using Hybrid Programming of visual basic6.0 and MATLAB7.0.1. Secondary network interface circuit is shown in Figure 5.

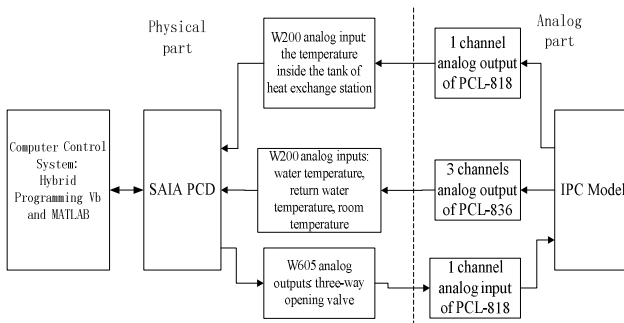


Fig. 5. Schematic diagram of interface circuit in secondary heating circuits

Experimental results show that, while Implementation steps of DMC-PID control process is about 10 steps or so, the control output reached to the pre-expectations and the system runs stably in the setting range. Meanwhile, the disturbance has little effect on system stability and thus the control effect is excellent. Compared with the previous survey results, the control effect can be achieved more rapidly and accurately.

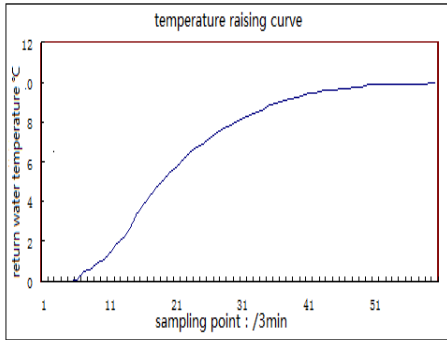


Fig. 6. The curve of unit impulse response

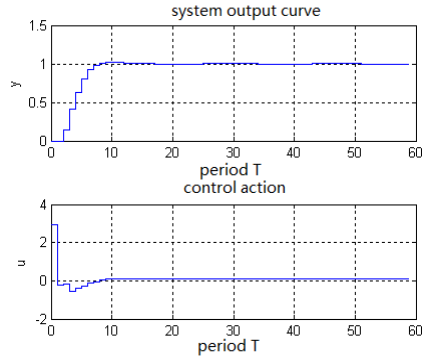


Fig. 7. DMC-PID operating process

6 Conclusion

This paper introduces the application of dynamic matrix control algorithm in gas boiler heating system. Compared with the existing methods, improved algorithm has better efficiency and thus has higher utility value.

References

1. Wei, C., Ting, Q., Wei, W.: Existent problems of gas boiler heating and application of energy saving techniques to boiler plants. *HeatingVentilating & Air Conditioning* 35(10), 104–110 (2005)
2. Qian, J.-X., Jun, Z.: *Predictive Control*. Chemical Industry Press, Beijing (2007)
3. Ding, B.-C.: *Predictive control theory and method*. Mechanical Industry Press, Beijing (2008)
4. Grimble, J., Ordys, A.W.: *Predictive control for industrial applications*. *Annual Reviews in Control* 25, 13–24 (2001)
5. Aufderheide, B., Wayne Bequette, B.: Extension of dynamic matrix control to multiple models. *Computers & Chemical Engineering* 27(8-9,15), 1079–1096 (2003)
6. Mahlia, T.M.I., Abdulmuin, M.Z.: Dynamic modeling and simulation of a palm wastes boiler. *Renewable Energy* 28(8), 1235–1256 (2003)
7. Xu, X.-W., Meng, X.-F.: Design of Hardware-in-the-Loop Simulation Universal Platform for Complex Measure and Control System. *Journal of System Simulation* (21), 126–132 (2009)

The Expert Controller for Dissolved Oxygen Control System in Wastewater Treatment

Jia Minzhi, Gao Binbin, and Jia Bo

College of Information Engineering, Taiyuan University of Technology, China
tutjzmz666@sina.com, {gaobinbin168, jiabo1551329}@163.com

Abstract. This article introduces a kind of expert controller which is aimed at dissolved oxygen control system in wastewater treatment of T-type oxidation ditch process. The use of the controller is to control the number, speed and running time of oxidation ditch's rotary brushes automatically. Meanwhile, it regulates the dissolved oxygen concentration of oxidation ditch. As a result; we can achieve the purpose of remove phosphorus from wastewater, denitrify, denitrify, and reduce the COD and BOD. At last, the effect of sewage disposal reaches the national emission standard.

Keywords: Controller, oxidation ditch, dissolved oxygen concentration, oxidation ditch's rotary brush.

1 Introduction

These sewage treatment plants, which use T-type oxidation ditch process, through control the number of oxidation ditch's rotary brushes or discs, speed and running time to regulate the dissolved oxygen concentration. Therefore, they can reach the aim of removing phosphorus from wastewater, denitrify, denitrify, reducing the COD and BOD, and make the sewage disposal reach the design requirement.

These dissolved oxygen concentration of sewage treatment plants influenced by many factors (like temperature, hydraulic retention time, sludge age, and the oxygen demand of various floras). It is a quite complicated process of physics and biochemistry. In addition, the setting of oxidation ditch is usually in the open air, and the change of the weather has a great impact on the dissolved oxygen concentration of oxidation ditch. Therefore it's difficult even impossible to build an optimized mathematical model to describe this process. So it is hard to adopt the algorithm of conventional PID controller or based on other control methods of mathematical model to achieve the demand of dissolved oxygen concentration. However, an experienced operator who controls the oxidation ditch's rotary brushes or discs reasonably can regulate the dissolved oxygen concentration greatly, so we design the following expert controller.

2 The Characteristics of Expert Controller [3][4]

2.1 The Description of Diversity Model

In the existing control theory, the design of control system only depends on the mathematical models of controlled plant. However, we adopt the expert controller in the technology of expert system, so it can dispose sorts of information no matter accurately or fuzzily, and it allows the model carrying out various forms of description.

The general model of expert controller can be shown is following below:

$$U=f(C, T, I)$$

In the equation, f is the intelligent operator; the basic form is shown below:

$$\text{IF } C \text{ AND } T \text{ THEN (IF } I \text{ THEN } U)$$

$C=\{C_1, C_2, \dots, C_m\}$ is the input set of controller; $T=\{t_1, t_2, \dots, t_n\}$ is the set of empirical data and fact set in the knowledge base; $I=\{i_1, i_2, \dots, i_p\}$ is the output set of inference engine; $U=\{u_1, u_2, \dots, u_q\}$ is the output of controller set.

The basic meaning of intelligent operator is that it depend on the input message C , empirical data of knowledge base T and rule to conclude, then depend on the deduced consequence I to output the corresponding of control behavior U .

2.2 The Flexibility of Online Processing

In the process of designing expert controller, the online information process and use is very important. The flexibility of online information process would improve the system's ability of process and the level of decision.

2.3 Flexible Control Strategy

It requires the controller who can adopt different forms of open loop and closed loop's strategy when industrial object have time-varying happen to itself or the site exist interference. Therefore through obtain the online information to flexibly modify the control strategy or controlled variable to ensure achieving better control performance.

2.4 The Hierarchical of Decision-Making Body

The design of intelligent controller should incarnate the principle of hierarchy whose kernel is mimicking human. That means depending on different levels of intelligence to consist of the decision-making with hierarchical structure.

2.5 The Real-Time of Inference and Decision

When we design an expert controller which is used for industrial engineering, in order to meet the demand of industrial engineering's real-time, the scale of knowledge base cannot be too big, and the inference engine should be as simple as possible.

3 The Determination of Expert Control Scheme

We took an incompact expert controller in a sewage treatment plant, which use rotary brush of oxidation ditch to regulate the dissolved oxygen concentration. After repeatedly research and experiment, by taking control of the rotary brush’s speed and running time, we achieved a comparisons ideal result of dissolving oxygen concentration, and the effect of wastewater treatment in sewage treatment plant was obvious improved. The basic thoughts of this scheme are:

1) The knowledge, inference mechanism and control rule mostly depend on the operators’ experience and the comprehensive knowledge of experts, so no need to build mathematical model of the objects. The structure chart is shown following figure1.

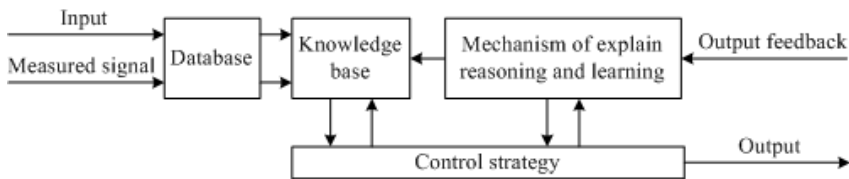


Fig. 1. The structure chart of expert controller

2) The controller algorithm is comparatively flexible and easy to achieve.

3) The experts’ intellectual is expressed by production rule in expert controller. Its basic structure is: [3][5]

IF <Condition> AND <Condition>
 THEN <Action> AND <Action>

4) Considering the experts’ infer by prior knowledge of control decision have indetermination of the certain degree. Therefore it can not be excessively nuanced for partitioning rule when adopt the production rule of production rule, otherwise it would reduce the robustness of control.

4 The Design Principles of Expert Controller

The input variables of this system are set as following: the dissolved oxygen concentration’s deviation of wastewater is $e(t)=r(t)-do(t)$, in this formula, $r(t)$ is the set of dissolved oxygen concentration(usually is 2mg/L), $do(t)$ is the actual measurement of dissolved oxygen concentration, the deviation of change in dissolved oxygen concentration is $\Delta e(t)=e(t)-e(t-1)$.The output variables include: the number of oxidation ditch’s rotary brushes, position , speed and running time. The destination of the control is to ensure the design requirements of dissolved oxygen concentration.

The database of system includes the information of deviation and change in deviation, process engineering, design requirement and climatic conditions. The rule base consist of operational approach by summarize from the operating experiences of

operator and knowledge of expert. And both of these constitute the uninterruptedly control for oxidation ditch, include the number of rotary brush, position, speed and running time, therefore it can ensure the diverse needs of dissolved oxygen concentration for different operation stages and achieve a better effect of wastewater treatment finally.

Depending on the process feature of a sewage treatment plant, in the specific climatic conditions, the operation mode of oxidation ditch's rotary brushes is: all of the rotary brushes will run to oxygenate with a high speed when the dissolved oxygen concentration of oxidation ditch is below 0.5mg/L. After the dissolved oxygen concentration have gone up 0.3mg/L, one of the rotary brushes will run with a low speed, then if gone up 0.5mg/L, the brush will stop running, and so on. When the dissolved oxygen concentration have reached 2mg/L, only two of these brushes will run with a low speed to mix the soil well with the water in the oxidation ditch. The dissolved oxygen concentration will continue rising even the system worked in low-duty, so the dissolved oxygen concentration of oxidation ditch gone up every 0.1mg/L, one of the brushes will stop, until all of these brushes stop running. After this condition last more then half an hour, all of the brushes will run 5 minutes to ensure the soil and the water mixed sufficiently, then all of the brushes will stop running, and so on. When the dissolved oxygen concentration of oxidation ditch begin to decrease, if it declined 0.3mg/L, one of the brushes will start running with a low speed, and when declined 0.5mg/L, the brush will run with a high speed, and so on. All of the rotary brushes will run with a high speed when the dissolved oxygen concentration is below 0.5mg/L. One point should be illustrate, when the dissolved oxygen concentration of oxidation ditch is enough and all of the brushes need stop running, the stop time should delay 100 seconds in order to avoid a frequently run and stop of the brushes while the dissolved oxygen concentration of oxidation ditch has tiny undulate.

5 The Realization of Expert Controller

Now we assume the number of rotary brushes which worked in high speed represented by RUN(h), the number of low speed represented by RUN(l), and the number of stop running represented by STOP, then the control law of the expert controller is shown below:

$$\text{IF } \Delta e(t) > 0 \text{ AND } e(t) \geq 1.5 \text{ THEN RUN}(h)=4 \text{ AND RUN}(l)=0 \text{ AND STOP}=0 \quad (1)$$

$$\text{IF } \Delta e(t) > 0 \text{ AND } 1.2 \leq e(t) < 1.5 \text{ THEN RUN}(h)=3 \text{ AND RUN}(l)=1 \text{ AND STOP}=0 \quad (2)$$

$$\text{IF } \Delta e(t) > 0 \text{ AND } 1 \leq e(t) < 1.2 \text{ THEN RUN}(h)=3 \text{ AND RUN}(l)=0 \text{ AND STOP}=1 \quad (3)$$

$$\text{IF } \Delta e(t) > 0 \text{ AND } 0.7 \leq e(t) < 1 \text{ THEN RUN}(h)=2 \text{ AND RUN}(l)=1 \text{ AND STOP}=1 \quad (4)$$

IF $\Delta e(t) > 0$ AND $0.5 \leq e(t) < 0.7$ THEN RUN(h)=2 AND RUN(l)=0 AND STOP=2 (5)

IF $\Delta e(t) > 0$ AND $0 \leq e(t) < 0.5$ THEN RUN(h)=1 AND RUN(l)=1 AND STOP=2 (6)

IF $\Delta e(t) > 0$ AND $-0.1 \leq e(t) < 0$ THEN RUN(h)=0 AND RUN(l)=2 AND STOP=2 (7)

IF $\Delta e(t) > 0$ AND $-0.2 \leq e(t) < -0.1$ THEN RUN(h)=0 AND RUN(l)=1 AND STOP=3 (8)

IF $\Delta e(t) > 0$ AND $e(t) < -0.2$ THEN RUN(h)=0 AND RUN(l)=0 AND STOP=4 (9)

IF $\Delta e(t) < 0$ AND $e(t) < 0$ THEN RUN(h)=0 AND RUN(l)=0 AND STOP=4 (10)

IF $\Delta e(t) < 0$ AND $0 \leq e(t) < 0.3$ THEN RUN(h)=0 AND RUN(l)=1 AND STOP=3 (11)

IF $\Delta e(t) < 0$ AND $0.3 \leq e(t) < 0.5$ THEN RUN(h)=1 AND RUN(l)=0 AND STOP=3 (12)

IF $\Delta e(t) < 0$ AND $0.5 \leq e(t) < 0.8$ THEN RUN(h)=1 AND RUN(l)=1 AND STOP=2 (13)

IF $\Delta e(t) < 0$ AND $0.8 \leq e(t) < 1$ THEN RUN(h)=2 AND RUN(l)=0 AND STOP=2 (14)

IF $\Delta e(t) < 0$ AND $1 \leq e(t) < 1.3$ THEN RUN(h)=2 AND RUN(l)=1 AND STOP=1 (15)

IF $\Delta e(t) < 0$ AND $1.3 \leq e(t) < 1.5$ THEN RUN(h)=3 AND RUN(l)=0 AND STOP=1 (16)

IF $\Delta e(t) < 0$ AND $1.5 \leq e(t)$ THEN RUN(h)=4 AND RUN(l)=0 AND STOP=0 (17)

6 Conclusion

We adopted the expert controller and time control to achieve automatic control the dissolved oxygen concentration of oxidation ditch in a sewage treatment plant, who adopts T-type oxidation ditch process in Shanxi province of China. The effect of control is good, and the sewage disposal of this sewage treatment plant reached the national standards of Grade II. Table 1 is the water's BOD₅ actual measured values compared with the values which calculated by the sewage treatment plant's model operating for one year. From the table we can get that the computational values of

model are consistent with the actual measured values and only in November and December have relatively large differences. This suggests that climate changes have relatively large impacts on the effects of wastewater treatment. Therefore in order to achieve an ideal effect of wastewater treatment, aiming at different kind of weather, we must adjust the parameters of expert controller and time parameters. So the effect of control will be better if we import self-learning mechanism of parameters into the expert controller, add the influence parameters of dissolved oxygen concentration into the database of controller, and take these parameters as basis for control decision.

Table 1. Data Comparison

Data(mouth)	Effluent BOD51 (mg/l)	Effluent BOD52 (mg/l)	Temperature (°C)
1	14.0	14.4	11.1
2	16.3	17.3	11.2
3	17.8	18.1	11.0
4	17.8	16.3	18.6
5	20.8	20.0	22.3
6	18.5	18.5	24.4
7	20.3	19.0	23.4
8	20.3	17.0	27.3
9	16.9	16.9	24.6
10	18.7	18.6	22.4
11	14.8	22.8	17.3
12	17.8	21.6	11.9

Note: BOD5 (mg/l) is the average of 5 days' biochemical oxygen demand, BOD51 (mg/l) is the computational values of model, BOD52 (mg/l) is the actual measured values.

References

1. Xingcan, Z., Yaxin, L.: Technologies for Nitrogen and Phosphorus Removal. China Architecture & Building Press, Beijing (1998)
2. Tingyao, G., Siqing, X., Zengyan, Z.: Summary on the Technologies of Biological Nitrogen and Phosphorus Removal from Municipal Wastewater. Chinese Journal of Environmental Science 20(1), 68–73 (1999)
3. Jingsun, F., Zixing, C.: Artificial Intelligence and Its Application. Tsinghua University Press, Beijing (1988)
4. Huaguang, Z., Xiangping, M.: Intelligent Control Theories and Their Applications. China Machine Press, Beijing (2005)
5. Qian, F., Mengxiao, W., Wei, T.: Aerobic treatment control system of APMP pulping wastewater based on expert system. China Pulp & Paper Industry 30(7), 87–89 (2009)

Distributed Fuzzy Control System for Cotton Cloth Bleaching

Xiaohong Guo and Congyan Yang

School of Science, Taiyuan University of Technology, West Yingze Steet 79,
030024 Taiyuan, China
{Guoxiaohong, Yangcongyan}@tyut.edu.cn

Abstract. Nowadays, under the progress of modern computer technology and intelligent control theory, the microprocessors centered distributed control systems provides a broad space for the development of textile industry. Among these systems, fuzzy control has become a major means of computer intelligent control. In this paper, in order to improve the production quality of the cotton cloth breaching process in the textile industry, a distributive fuzzy control system for cotton cloth bleaching is implemented with the industrial control computers, the variables, the rules and the controlling process of the fuzzy controller in this system are designed after analyzing the operational sequence of cotton cloth bleaching. The practice application showed that this system turns out to be meeting the design requirements with a 95% average approved quality of cotton cloth.

Keywords: Distributed control system, Fuzzy control, Industrial control computer, Cotton cloth bleaching, Textile industry.

1 Introduction

The bleaching of cotton cloth, the instinctive quality of cotton for printing and dyeing processing, is a key process in the textile industry [1]. The quality of this process has a decisive influence on the quality of the finished cloth. Under the progresses of the modern computer technology and the intelligent control theory, the traditional control mechanisms of the cotton cloth bleaching is being gradually substituted by modern microprocessors centered distributed control systems, providing a broader space for the development of the textile industry [2, 3].

Fuzzy control is an intelligent control method based on fuzzy set theory, fuzzy language variables and fuzzy logic reasoning, and is suitable for measuring the imprecise controlled object or the objects with dynamic changes [4]. Owing to the characteristics of robustness, nowadays fuzzy control has become a major means of the computer intelligent control [5]. In this paper, we developed a distributive fuzzy control system with the industrial control computers for the automatic control of cotton cloth bleaching within a workshop, and designed the fuzzy variables and fuzzy rules of the fuzzy controller in this system after analyzing the operational sequence of

cotton cloth bleaching. The practice application shows that this system turns out to be meeting the design requirements with a 95% average approved quality of cotton cloth.

This paper is organized as follows. Section 2 introduces the distributive fuzzy control system for cotton cloth bleaching control and the key subsystems. Section 3 analyzes the operational sequence of cotton cloth bleaching, and categories the stages in the sequence into four groups. Section 4 devises the fuzzy controller, designs the fuzzy variables and fuzzy rules, and gives an applicable controlling process for cotton cloth bleaching. At last, Section 5 concludes.

2 Distributive Fuzzy Control System for Cotton Cloth Bleaching

Figure 1 depicts the structure of the distributed fuzzy control system for cotton cloth bleaching. The system consists of five subsystems, namely the data collecting subsystem, the fuzzy controller subsystem, the dynamic displaying subsystem, the failure alarming subsystem, and the communicating interface subsystem. The functionalities of these subsystems are as follows: (1) Data collecting subsystem: the key data in each cycle of the bleaching process is collected by detectors in this subsystem before being sent to the computer control center for analysis and calculating. Then, they are stored and transmitted to given controllers. (2) Fuzzy controller subsystem: controls the operational sequences of cotton cloth bleaching according to the standard technical requirements as well as the fuzzy control algorithms. (3) Dynamic displaying subsystem: the collected data is also sent to a large display screen through which the operators can supervise the constantly changing data. (4) Failure alarming subsystem: the alarm will be triggered when something abnormal is detected in the bleaching processes or the collected data varies too far from the required standard.. At the same time, the reason for the variance will appear on the displaying screen. (5) Communicating interface subsystem: though this interface the control center can exchange information with the upper management systems in the company.

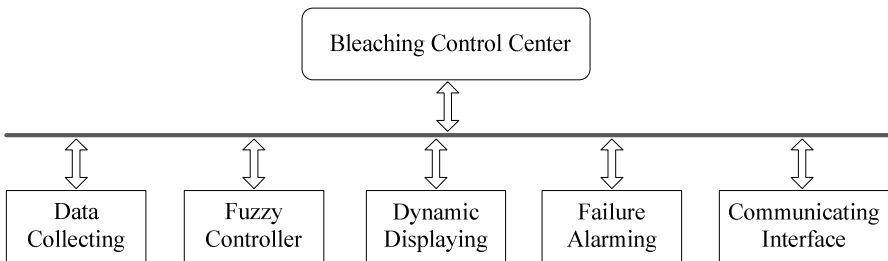


Fig. 1. The structure of the distributive fuzzy control system for cotton cloth bleaching

3 The Operational Sequence for Cotton Cloth Bleaching

The technology used in the bleaching process determines the quality of the final product. The bleaching technology used in this system is known as “Chlorine and

Oxygen Double Bleaching". It means that the control system operates through the interaction of both Chlorine and Oxygen to produce the bleached cloth. The bleaching agents are HClO and H₂O₂. Through the interaction of these two oxidizing agents in sufficient concentration and with the presence of heat, the ideal bleach is attained.

In detail, the operational sequence of cotton cloth bleaching process is illustrated in Figure 2. Among these stages, the data for the concentration of the bleaching liquid is collected and controlled during the three bleaching stages, i.e., two Chlorine Bleaching stages plus an Oxygen Bleaching stage, and the Sulfuric Acid wash stage; the data of time is collected and controlled during the rest stages, whereas the data of temperature was collected and controlled during the steaming stage.

Water Washing → Chlorine Bleaching I → Rest → Chlorine Bleaching II → Rest → Sulfuric Acid Washing → Water Washing → Oxygen Bleaching → Steaming → Rest → Water Washing

Fig. 2. Stages in the operational sequence of cotton cloth bleaching process

In Figure 2, the underlined stages are crucial ones in the operational sequence of cotton cloth bleaching. They are independent to each others. Therefore, we can divide these key stages into four controlling groups, as illustrated in Table 1.

Table 1. Four controlling groups in the operational sequence of cotton cloth bleaching process

Group Name	Grouped Stages	Controlling Variables
Chlorine Bleaching I	Chlorine Bleaching I, Reset	Concentration, time
Chlorine Bleaching II	Chlorine Bleaching II, Reset	Concentration, time
Sulfuric Acid Washing	Sulfuric Acid Washing	Concentration, time
Oxygen Bleaching & Steaming	Oxygen Bleaching, Steaming, Rest	Concentration, time, temperature

As Table 1 shown, except for the Sulfuric Acid Washing group, every group has a Bleaching stage and a Rest stage. In the Oxygen Bleaching & Streaming group, there is one more stage, i.e., the Steam stage. Each group has its own standard demands which can be controlled separately. When every stages works well according to these standards, the ideal bleached cotton cloth will be produced.

4 Design of the Fuzzy Controller for Cotton Cloth Bleaching

Based on Table 1, we devise the fuzzy controller, the key subsystem of the distributive fuzzy control system for cotton cloth bleaching, with four control units, as shown in Figure 3. Every control unit corresponds to a group in Table 1, and can be controlled independently. According to this structure, we will design the fuzzy variables and the fuzzy rules, and sketch the applicable controlling process for the fuzzy controller in the following subsections.

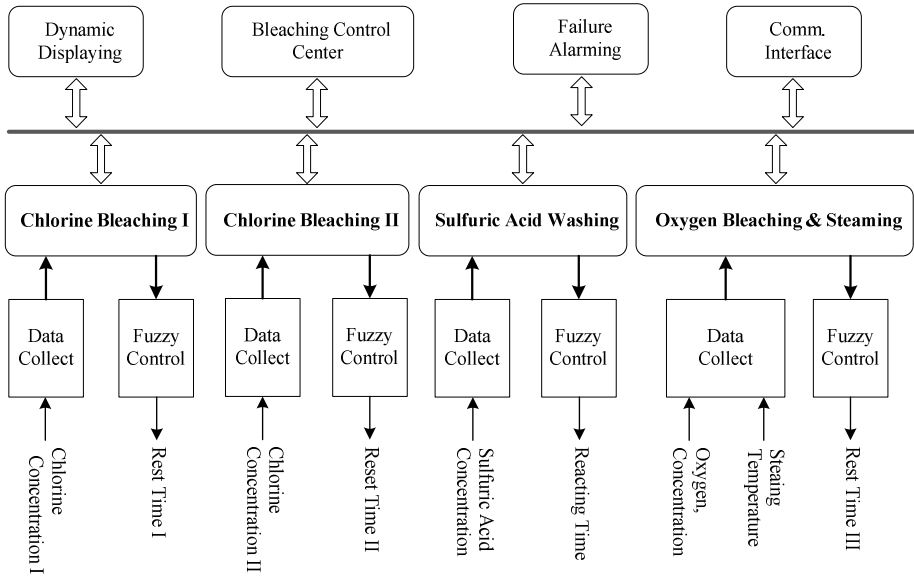


Fig. 3. The structure of the fuzzy controller with four control units

4.1 Design of Fuzzy Variables

The fuzzy controller has five inputs and four outputs. We define the fuzzy sets of these variables as in Table 2. According to Table 1, these variable can also be categorized into four groups. Taking the Oxygen Bleaching & Streaming group as an example, the related input variables are CO (Oxygen Concentration) and ST (Steaming Temperature), the related output variable is RT3 (Rest Time 3), with fuzzy membership functions shown in Figure 4.

Table 2. Fuzzy Sets of Input/Output Variables

	Variables	Fuzzy Sets
Input	CC1 (Chlorine Concentration I)	{low, middle, high}
	CC2(Chlorine Concentration II)	{low, middle, high}
	SAC (Sulfuric Acid Concentration)	{low, middle, high}
	CO (Oxygen Concentration)	{low, middle, high}
	ST (Steaming Temperature)	{low, middle, high}
Output	RT1 (Rest Time I)	{short, middle, long}
	RT2(Rest Time II)	{short, middle, long}
	AT (React Time)	{short, middle, long}
	RT3 (Rest Time III)	{short, middle, long}

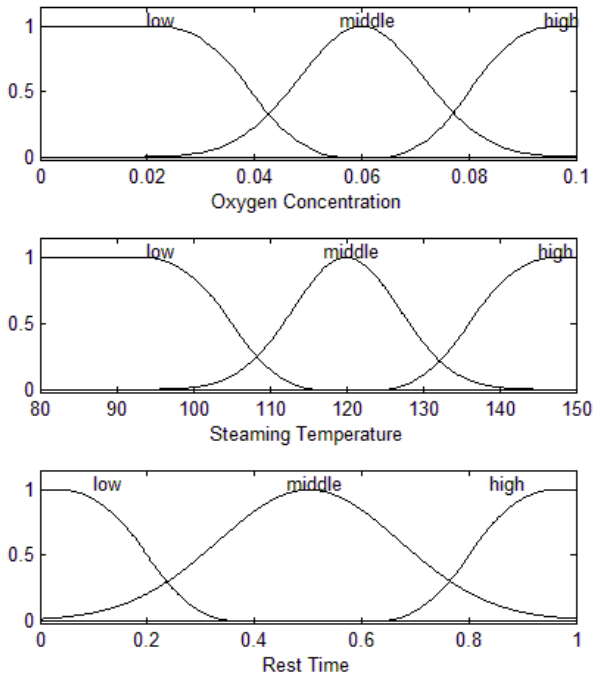


Fig. 4. Graph of sample Fuzzy membership functions

4.2 Design of Fuzzy Rules

According to the fuzzy optimal control strategy, we designed the fuzzy rules of each group in table 1 after careful experiments and abundant practice. Among these rules, the one with regard to the Oxygen Bleaching & Streaming group is most complex, and they are illustrated in Table 3.

Table 3. Fuzzy Rules w. r. t. Bleaching & Streaming group

RT3		ST		
		low	middle	high
OC	low	long	long	long
	middle	long	middle	middle
	high	middle	short	short

Lastly, since the output of the fuzzy controller is a fuzzy value, a proper defuzzification method is necessary to transform the fuzzy value into a precise one. After experiments and practice, we choose weight average method for defuzzification as

$$T = \sum \beta_i k_i / \sum \beta_i \tag{1}$$

With fuzzy rules in Table 3 and formula (1), we can predict a proper rest time or reacting time according to given input values, e.g., the Chlorine concentrations, the Sulfuric acid concentration, the Oxygen concentration, as well as the Steaming temperature. Taking the Oxygen Bleaching & Streaming group as an example, the graph of the output RT3 to the inputs OC, ST are shown in Figure 5.

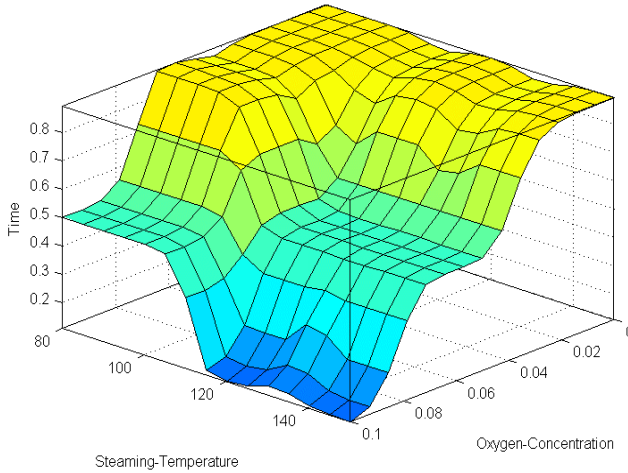


Fig. 5. Graph of the output RT3 to the inputs OC, ST w.r.t. the fuzzy rules of the Oxygen Bleaching & Streaming group

4.3 Design of Applicable Controlling Process

In this distributed fuzzy control system, the fuzzy controllers are built up with industrial PC computers, the control center is a super computer, and all the detectors are strictly following the industrial standards. What's more, the controlling process of the fuzzy controller is shown in Figure 6, and it can be operated either automatically or manually.

5 Conclusions

In this paper, we implemented a distributive fuzzy control system for automatic control of the bleaching process of cotton cloth in textile industry. As the centre of this system, the fuzzy controller is devised and its fuzzy variables, fuzzy rules, controlling process are designed accordingly, after a careful analysis of the operational sequence of cotton cloth bleaching. Till now, this system has been successfully implemented and well applied. After adjusted application and tests for several months, this system turns out to meet the design requirements, with a 95% average approved quality of cotton cloth. For this reason, it's awarded as the second prize in the Shanxi Provincial Science Progress Prize, and has been weighed heavily in favor of the Textile Industries of Shanxi.

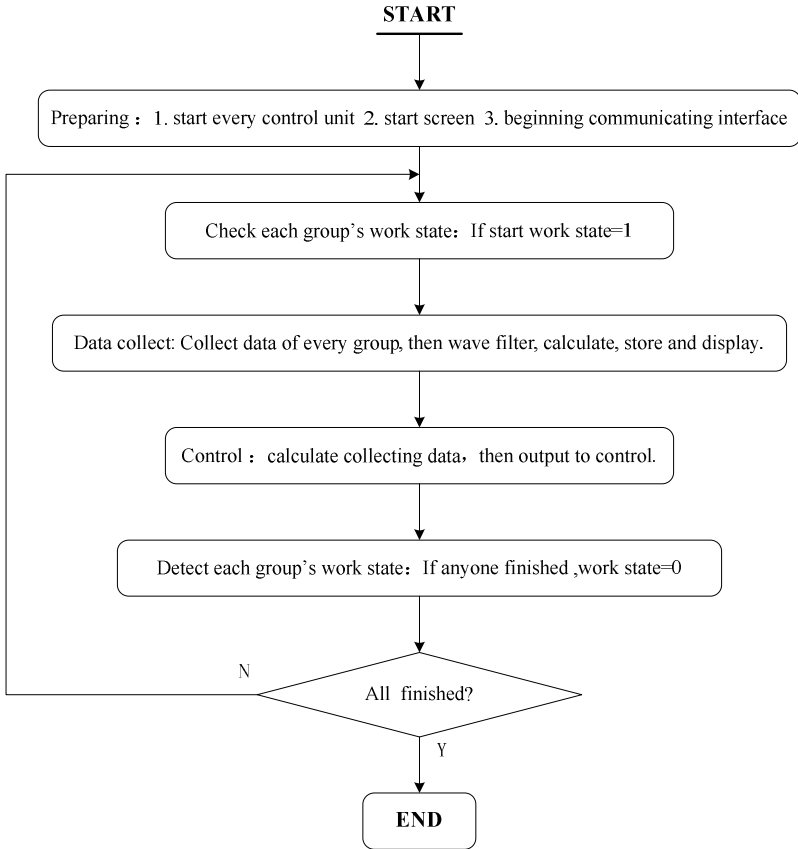


Fig. 6. The controlling process of the fuzzy controller

References

1. Powell, L.: Industry Applications. *IEEE Transactions on Control System Technology* 18, 176–185 (2004)
2. Berustein, D.S.: Control system. *IEEE Transactions on Control System Technology* 12, 201–220 (2003)
3. Valanvanis, K.P.: Robotics and Automation. *IEEE Transactions on Robotics and Automation* 22, 134–145 (2004)
4. Huiling, C.: Intelligent Control and Its Applications in Industrial Process Control. In: *The 7th International Symposium on Test and Measurement*. IEEE Press, Los Alamitos (2007)
5. Passino, K.M., Yurkovich, S.: *Fuzzy Control*, 1st edn. Addison Wesley Publishing Company, Reading (1997)

A Way to Diagnose the Rolling Bearing Fault Dealt with Wavelet-Packet and EMD

Xiao-Feng Liu, Shu-Hua Wang, Yong-Wei Lv, and Xuan Lin

College of Electrical and Power Engineering, Taiyuan University of Technology, West Yingze Street No.79, 030024 Taiyuan, Shanxi, China
tianmuqin@tyut.edu.cn

Abstract. The way to diagnose the rolling bearing fault in advance is a key to safe production and avoids serious accidents in technology. This paper puts forward a update way to diagnose the fault in which the wavelet deletes the noises by the original signals handed out and differentiates the originals by frequency, and then use EMD to resolve the low-frequency signals got by wavelet decomposition and reconstruction to get a number of inherent IMF, each function of which is analyzed by time-frequency to know the fault frequency from spectrogram and compare Fourier transform and wavelet transform with strengths and weakness of the way to diagnose. This experiment has shown by the diagnosis we can pick up the fault frequency effectually and easy to judge and diffentiate fault types.

Keywords: wavelet, EMD, IMF, bearing fault diagnosis.

1 Introduction

With the development science and technology, mechanical equipment has become more complicated, mechanical failure has happened more frequently to result in serious consequence, even great economic loss. Rolling bearing is the most popular mechanical spare parts used in electric power, neurochemistry, smelting, machinery, aircraft industry, and it is also one of the components easiest to damage. It is counted that among rolling machinery equipment which use rolling bearing, about 30 percent mechanical fault is related to rolling bearing damage. Therefore, the analysis of diagnosis of rolling bearing fault is particularly important in production. There are frequency domain analysis and time domain analysis in the traditional diagnosis of rolling bearing fault. They have good effect on rolling bearing fault, but as for local defeat, the application of the diagnosis is not satisfied, especially in the earlier period of defeat. In recent years, the methods to diagnose exist constantly, such as time-frequency analysis which can be applied to non-stationary signal analysis effectually. Spectrum analysis which makes up traditional Fast Fourier Transform adapts only to the fault of stationary signals. Wavelet-pocket transform has improved middle and high frequency resolution or decomposition .Overcome the fault in short-time FFT and wavelet Transform. Wavelet-pocket transform has advantage of drawing bearing fault. However current signal analysis techniques haven't made great progress in

picking up the feature of vibration signals of low-signal noise ratio. During rolling bearing working, drawing vibration signals is early, interfered with a great deal of non-tested parts' vibration to cause effective signals to disappear, especially rolling bearing vibration signals which go through complex paths, always easy to have fault information overwhelmed in the background noises and interference, and have it difficult to draw signal[1].

This paper takes rolling bearing as a research object, analyses the data of vibration signals in the fault state of rolling bearing, adopting a new stationary signal analysis--Empirical Mode Decomposition (EMD):This method can show the data of vibration signals obviously from local signals of rolling bearing in the fault, and pick up them effectually, overcome the restriction of FFT, and pick up them effectually, overcome the restriction of FFT, and so supply a fresh way for analyzing the vibration signals of rolling bearing.

2 The Application of the Wavelet Transform in Rolling Fault

Unavoidably, there existed noise in our experiment. When the noise was too loud or signals themselves were complex, the following signals were dealt not happily. Wavelet analysis was time-frequency analysis recently applied more widely. Because it had frequency window when the local part between time domain and frequency domain was changing .It is superior to traditional Fourier analysis non-stationary signals. Wavelet analysis classify signal frequency, raise signal resolution power in time domain and frequency domain, and reduce and weaken the other signals and noise [2].

In the process of handling signals, we adopted a wavelet basis function ψ to remove original signal noises, and the signals handed out were decomposed by wavelet and at four or five levels, low-frequency was restricted, and only to analysis the low-frequency further.

The wavelet decomposition contracture of signals shows as a tree. Here take three levels decomposition as an example to show that the wavelet decomposition tree is shown as Figure 1: a stands for low-frequency, d stands for high-frequency, the final order number shows the number of wavelet decomposition levels (that is scale).Decomposition shows as following [3]:

$$S=aaa_3+daa_3+ddd_3+dad_3+ddd_3. \tag{1}$$

There are many forms of decomposition, the real process of dealing is generally based on signal energy to decide further decomposition strategy. After choosing a wavelet, we took advantage of the principle above to decompose the signals at N level wavelets, and choose a threshold obtained from the each level number got from the decomposition to carry out handling soft threshold to avoid discontinuity. Finally, we used the coefficient handled reconstruct and recover the original signal by wavelet-pocket to realize removing noises.

3 The Empirical Mode Decomposition (EMD)

EMD was put forward by Norden. E.Huang, a American-Chinese scientist in recent years. EMD is a new way to deal with signals, in which signals are decomposed into a

certain number of inherent Intrinsic Mode Function (IMF) and the other linear sum, according to the feature of the signals and time scale. This IMF reflects the internal feature of signals, and others show signal's trend.

Suppose a EMD: Any signal is made up of different inherently simple vibration mode, each having the same amount of extreme point and zero cross point ,and there being only an extreme point between two close cross point, and any two modes are independent. In this way, any signal can be decomposed into a number of the sum of the inherent internal mode function. However, it has to meet two conditions:

- 1) The number of the IMF extreme by decomposition, including maximum and minimum, and the number of passing zero point must be equal or at most one less.
- 2) At any time point, local average number of upper envelope defined by signals, local maximum and lower envelope defined by local minimum is equal to zero.

Inherent internal mode function can be got by the following ways:

- 1) Find out the local extreme point of a signal $x(t)$.
- 2) By connecting all the local maximum with cubs spine, upper envelope $e+(t)$ got, and lower envelope $e-(t)$ got, too, local average number $m(t)$ is calculated by Format (1):

$$M(t)=(e+(t)+ e-(t))/2. \tag{2}$$

- 3) Calculate Difference function $2i(t)$:

$$2i(t)=x(t)-m(t). \tag{3}$$

Check up whether $z_i(t)$ meets the condition of IMF .If not, take $z_i(t)$ as a new number to deal with. Do these above over and over again, if $z_i(t)$ meets the need, $z_i(t)$ is the first IMF, write it as $y_1(t)$.

Separate $y_1(t)$ from $x(t)$, therefore, EMD can decompose any signal $x(t)$ into the sum of k basic mode number and a redundancy $x_k(t):x_1(t), x_2(t), \dots, x_k(t)$, includes composition from high to low frequency signals, but not Equal width. Therefore EMD is a suitable signal decomposition $x_n(t)$ shows the signal trend[4].

Fault frequency is calculated as following (4):

Inner ring fault frequency:

$$F_i = 0.5Z(1 + d \cos \alpha / D) f \tag{7}$$

Outer ring fault frequency:

$$F_o = 0.5Z(1 - d \cos \alpha / D) f \tag{8}$$

Rolling ball fault frequency:

$$F_b = 0.5D[1 - (d \cos \alpha / D)^2] f / d \tag{9}$$

The model of the experiment bearing is 4309, the number of bearing is 8.

3.1 Experiment Methods and Steps

- 1) Use wavelet-packet to remove noises of vibration signal x , and FFT transform signals x_c handed out – original signals x to get the spectrogram.
- 2) The de-noising signals x_c are decomposed by wavelet and reconstructed to get low-frequency signals, which is transformed by FFT to get spectrogram.
- 3) Use the way EMD to decompose signal y to get a number of IMF function, and FFT transforms each IMF function to draw fault frequency.
- 4) Contrast each spectrogram and analyze it to draw a conclusion.

3.2 Analysis of the Fault of Inner Ring

A series of figures can be got by dealing with signals according to experiment steps. The low-frequency signals a_5 got by four-level decomposing and reconstructing by wavelet only include low-frequency signals as Fig.1. FFT transforming the signals shows that their frequency composition is too complex to show fault frequency clearly.

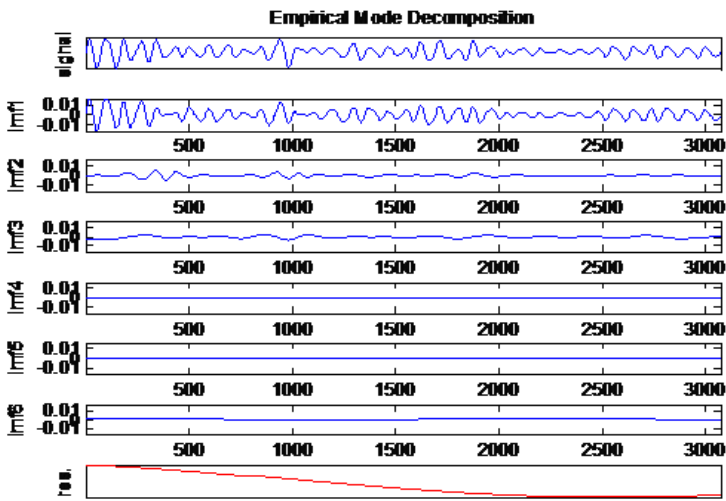


Fig. 1. The decomposition of a_5 by EMD

A number of IMF functions can be got by EMD decomposing low-frequency reconstruction signals. IMF functions lower from the top frequency to the bottom in turn, and local signal frequency is obvious, which is helpful to draw the fault frequency, and FFT transforms each IMF function. Spectrogram got from IMF 1 can show signals 119Hz clearly, and it is close to the two times inner ring fault frequency (118.8Hz) as Fig.2.

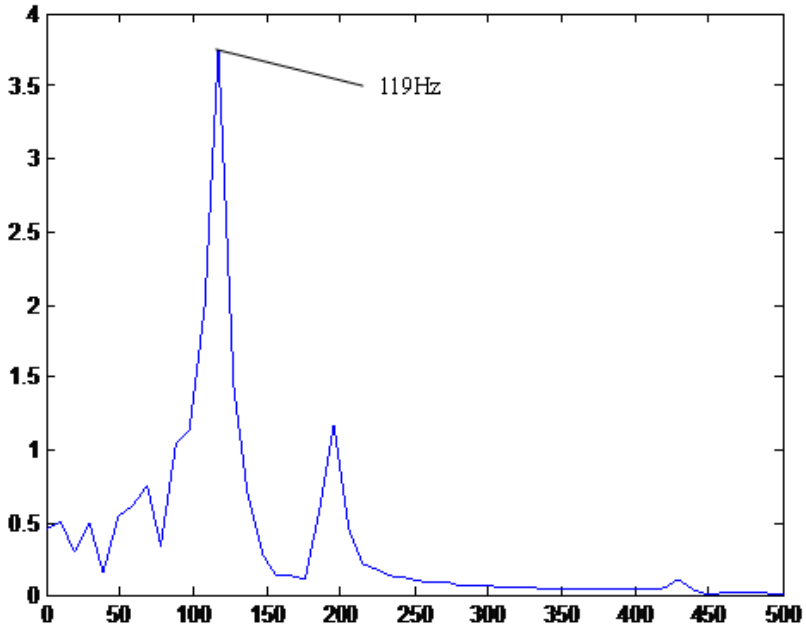


Fig. 2. The spectrogram of IMF1

Spectrogram got from IMF 2 shows signals of 59Hz is especially clearly, and close to the inner ring 59.23Hz fault frequency handed out by calculation, which shows that there is the fault of inner ring 3. The experiment and analysis of rolling ball as Fig.3.

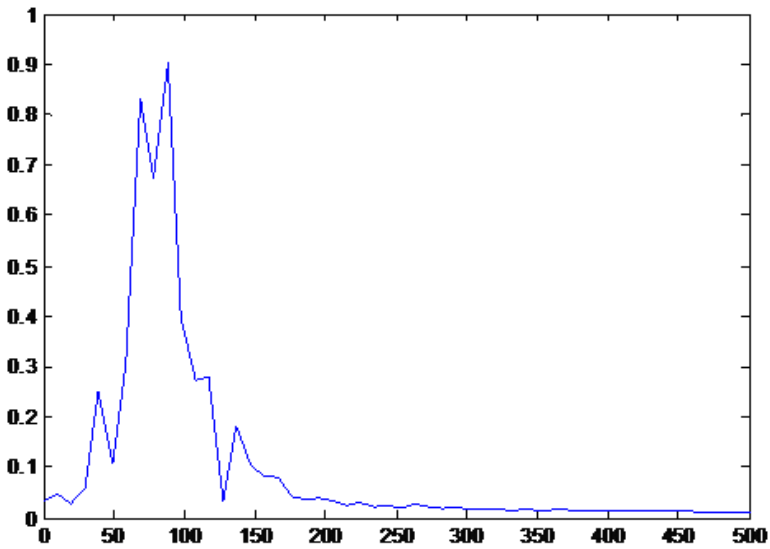


Fig. 3. The spectrogram of IME2

In order to test the affectivity of this method further, the fault bearing If the rolling ball was at the speed of the 710r/min. the figure shows the de-noising by wavelet-pocket is very good, and the decomposing and reconstructing function of wavelet-pocket are easy to analyze later, However, these methods are still restricted .

Rolling ball's fault frequency is 10.34Hz, and close to its frequency of calculation, and also clear at its frequency multiple, three-time frequency, four-time frequency, which shows the fault appears in the rolling ball. Analyzing the fault of inner ring, the turning frequency of the bearing is 10Hz, which is the same as the fault frequency of the rolling ball analyzed, and therefore, there is no turning-frequency composition. These show that DMD decomposition is an effective method to draw the information of the fault in the fault in the earlier time.

4 Conclusion

A series of experiment and analysis above show that FFT time-frequency transform analysis is not adapt to deal with bearing vibration signals and non-stationary signals. Contrast the spectrum analysis of bearing fault signals by the wavelet-pocket decomposition with original spectrum can show the useless high-frequency interfere signals removed .This explains the wavelet-pocket decomposition of vibration signals has de-noising and location ability ,but not able to define the feature frequency of weak or small fault , and even not to judge the types of bearing fault .The similar signals, got by EMD decomposing the wavelet -pocket are further decomposed t get a number of IMF function . The spectrogram of IMF function can show the fault feature frequency so as to define fault type. This method to combine the wavelet-pocket decomposition with EMD can separate the weak fault feature frequency with cow-frequency and overcome the restriction of traditional methods .The experiment proved that adopting this method to diagnose fault bearing can separate signals effectually and get beating fault feature information .The application of wavelet-pocket and EMD supplies a new method for bearing fault in the early period.

References

1. Yi, T., Liang, C., Round, Z.: Cepstrum-based Fault Diagnosis of Bearing Technology. Machine Tools and Hydraulic (2009)
2. Chen, L., Yan, X., Holly, X.: EMD flash visual evoked potential based on the single extraction method. Southern Yangtze University, Natural Science (2008)
3. Wang, L., Ming, W., Qinghai, Z.: Corona discharge radiation signal analysis, the selection of wavelet basis function. Ordnance Engineering College 18(3), 11–14 (2006)
4. Zhou: MATLAB Wavelet Analysis Advanced Technology. Xidian University Press, Xi'an (2006)

Exploring of Intelligent System for Oxygen Respirator Intelligent Training Check System Used in Colliery Salvation

Mu-Qin Tian, Sen-Sheng Wei, Mu-Ling Tian, and Jun Guo

College of Electrical and Power Engineering, Taiyuan University of Technology,
West Yingze Street No.79, 030024 Taiyuan, Shanxi, China
tianmuqin@tyut.edu.cn

Abstract. Oxygen respirator intelligent training check equipment used in colliery salvation is an intelligent system combining analysis, control and monitoring in performance, which can analyze the gas in the system, check the change of pressure, exhalation resistance and inhalation resistance, examine both the inhalation temperature and the environment temperature, automatically or manually change the treadmill speed, grade and time to control the work of ambulance man. It changes the traditional training model for ambulance man, makes check quality and quantity. Everything is based on data and eliminates the inkling for artificial feeling. This will realize ambulance man body information collection truly, improve the training mode to promote team members' diathesis and supply science gist to judge team members' body condition.

Keywords: Oxygen respirator; intelligent training; colliery salvation.

1 Introduction

This system collects analysis, monitoring and control together and reforms the traditional training module. It can check main parameters of instrument lively while team members wearing the respirators and records the physiological effect of ambulance members so as to offer science and reasonable evidence for member physical quality judge and selection.

2 System Function

Whole system includes information maintenance, speed control maintenance, self judgment, check management, check search, check, data base deal, history data deal, password management, equipment person management, and so on.

Information maintenance, which relates to some equipment capabilities' essential maintenance, such as measuring rage, precision, and so on, was done by the deviser Password is set here to prevent user change. This include basic parameter maintain and speed instruction character maintain.

Basic parameter maintain has the task to demarcate and calibrate the measure range of the instruments used in the equipment, judge, correct and fine adjust the change of output current and linearity of sensors (differential pressure transmitter, temperature sensor, mass flow meter, oxygen analyzer and CO₂ analyzer) in intelligent training check equipment. This will make the error adjustment and correction because of sensors or instruments replacement during the use convenient Standards will be put forward for the equipment check in the system. The time sequence maintain uses the same speed level and running time as treadmill's auto control time before leaving factory.

Speed control symbol is the control coding of signal module corresponding to treadmill's every speed level. When one way is out of work, code change is not need, but just input the changed control coding for the hardware way change.

Speed sequence maintain is the default treadmill speed level and running time before leaving factory. Since users always wish individually setting speed and time, user speed control maintain is designed and the treadmill works according this while auto control.

After system maintain, both static and dynamic tests can be done. For the correct test results, normal condition of equipment must be guaranteed firstly. So, hardware of the whole test system must make self diagnose both before the test and during the test. Cues will be shown according to whether the checked parameters such as air tightness under low pressure, exhaust pressure, self supply pressure, fix quantify flux, auto supply great flux and manual supply great flux are in the permission range. During the test, all signals in the system are checked. If abnormal signals are found, voice alarm will be given.

The tests include dynamic tests (function I) and static tests(function II). Manual speed control, leaving factory speed control and user speed control are included in dynamic tests.

The difference among manual speed control , leaving factory speed control and user speed control lies in whether the speed of treadmill is adjusted manually or use the speed and time while the equipment leaves factory, or use the speed and time according to the user. According to the sampling period, the whole time will be 240 min. (sampling period equal to 10 min, meaning one point record every 10 min) and

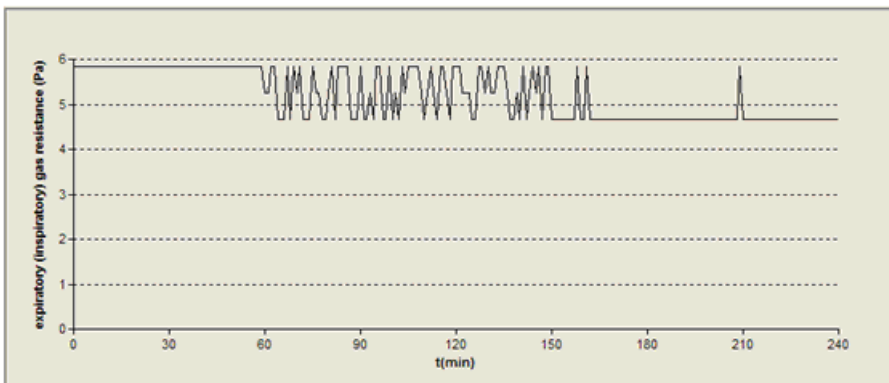


Fig. 1. Expiratory (inspiratory) gas resistance(Pa)

120min (sampling period equal to 5 min, meaning one point record every 5 min). Curves show one point every 10 second. There is a rapid test module in manual speed control, meaning whole test will take only 24 min. The test data include speed, inspiration resistance, expiration resistance, CO₂ content, O₂ content, inspiration temperature, environment temperature, and so on. All these will be print in table.

Moreover, curve of every physical quantity can be shown separately or assembly to make analyze and compare convenient. If computer does not work for power off, system will work from the point before interruption and keep the relative parameters connective. Pause is also permitted, knock 'test on' will continue the test. A dynamic test process can be seen in Fig. 1.

Static test (function II) will check the positive pressure airtightness, exhaust pressure, self supply pressure, fix quantify flux, auto supply great flux and manual supply great flux. Before test switch is on "auto" position, then connect the respirator and the test equipment. Knocking the test item in the computer, system will give a hint about the connect of item and relative hardware. If connection is right, a series of chaining control signals will be given and the relative control phase will be shown through signal lamps. With the test, change of the physics quantity will be shown. Whether the result is eligibility or not can be judged according to the standards, and the result can be shown both in table and in sound. All result will be saved to database and printed.

After test, the computer will show the test interfaces, which include test data display and test items selection. The eligibility ranges are given during data display for comparing. While connected to the respirator, function II is ready to begin. The following are the contents:

1) When test low pressure airtightness, mandrill should be plug in the hole on the lower shell of the respirator, expiration tube and inspiration tube are connected to the expiration tie-in and inspiration tie-in separately which lie in the left of the test equipment. Knocking 'low pressure airtightness', system will add pressure to the respiration until the default value is obtained. After keeping 20 second, begin timing. One minute later, computer will show the test value and audio. If there is disqualification, acousto-optic alarm will be sent out.

2) When test exhaust pressure, mandrill should be plug out, expiration tube and inspiration tube are connected to the expiration tie-in and inspiration tie-in separately which lie in the left of the test equipment. Knocking "exhaust pressure", system will remind "please remove the mandrill". After being confirmed, system will add pressure to the respiration until it reaches exhaust point and keep stabilization. Computer will show the test value and audio. If there is disqualification, acousto-optic alarm will be sent out.

3) While test self supply pressure, mandrill should be plug out, expiration tube and inspiration tube are connected to the expiration tie-in and inspiration tie-in separately which lie in the left of the test equipment. Open the gas bottle. Knock "self supply pressure". System will remind "please open the gas bottle." After being confirmed, system will evacuate until it reaches self supply point and self supply begin. While self supply pressure becomes stabilized, computer will show the test value and audio. If there is disqualification, acousto-optic alarm will be sent out.

4) When test fix quantify flux, open the breath cover of the respiration, connect the fix quantify flux tie-in to the adjustable valve, open the bottle, knock 'fix quantify flux', system will remind "please plug in tongue blade, open the bottle". After being confirmed, when flux becomes stabilized, computer will show the test value and audio. If there is disqualification, acousto-optic alarm will be sent out.

5) While test auto supply great flux, inspiration tube should be connected to auto supply great flux tie-in. Turn the auto supply/manual supply switch on the left corner to manual supply. Open the gas bottle. Knock "auto supply great flux". System will remind "please turn the great flux to 120 L/min, open the bottle". After being confirmed, vacuum pump work simultaneity. Adjust regulating knob to make flux reach request and stop adjusting. When flux becomes stabilized, computer will show the test value and audio. If there is disqualification, acousto-optic alarm will be sent out.

6) When test manual supply great flux, turn the two-way switch nut to the manual supply great flux tie-in. manual supply tube of the respirator will detach the manual supply tie-in of the respiration house, but connect to special tube. Another end of the tube should be connected to the switch nut and the expiration tube of the respiration should be connected to the test equipment. Turn the auto supply/manual supply switch on the left corner to manual supply, open the bottle, knock "manual supply great flux", press "manual supply valve" of the respiration. When flux becomes stabilized, computer will show the test value and audio. If there is disqualification, acousto-optic alarm will be sent out.

That is Process for pressure leak detection:

```
wac=1
timer(0)
timer(30)
if wac=1 and delay60=0 then
    if delay601=0 then
        a22=a21
        sle_3.text=string(a22,"###0")+ "(Pa) "
    end if
    timer(0)
    timer(1)
    delay601=delay601+1
    sle_2.visible=true
    sle_2.text=" Timing of 60sec for pressure leak
        detect : "+string(delay601)+"sec"
    if delay601=60 then
        delay60=1
```

```

        end if
    else
        if delay60=1 then
            timer(0)
            sle_2.visible=false
            sle_2.text=""
            a33=a22 - a21
            if a33<0 then
                a33=0
            end if
            UPDATE srcssj
            SET cssj =:a33
            WHERE srcssj.csxm =:csmc1
            using sqlca;
            dw_1.retrieve()
            SELECT srcssj.sx,srcssj.xx
            INTO :a221,:a222
            FROM srcssj
            WHERE srcssj.csxm = :csmc1
            using sqlca;
            if (a33>=a222) and (a33<=a221) then
                UPDATE srcssjzt
                SET cszt = :cs1, cdck = :cs2
                WHERE srcssjzt.csxm = :csmc1
                using sqlca;
                sle_4.text=string((a21),"####0")+ "(Pa) "
                as_filename=" Pressure leak detection
Qualified.wav"
                Function
                PlaySound(as_filename,ai_option)
            else
                UPDATE srcssjzt

```

```

        SET cszt = :cs1, cdck = :cs3
        WHERE srcssjzt.csxm = :csmc1
        using sqlca;
            a1="#051001"+char(13) // alarm
            ole_1.object.output=a1
        cb_4.TriggerEvent(Clicked!)
        sle_4.text=string((a21), "###0")+ "(Pa) "

        as_filename=" Pressure leak detection
Unqualified.wav"

        Function
PlaySound(as_filename, ai_option)
            end if
            dw_2.retrieve()
            a1="#061201"+char(13)
            ole_1.object.output=a1
            cb_4.TriggerEvent(Clicked!)
        end if
    end if

```

The results can be inquired no matter they come from dynamic (function I) test or static (function II) test. Inquiring can be down according to the combined conditions or only dynamic inquiring. What is more convenient is that if the results of dynamic(function I) or static(function II) test are not printed for some reasons, another printing can be done. System will print the data in the type of table or curve as the inquire condition. This system also adds some test modules to make test and demarcate convenient, so as to resolve the inconvenience between hardware person and software programmers.

For safety and convenient data displace, system has database backup, and can release and delete history data to avoid repeated record of system maintain data.

Otherwise, system also builds files for equipment and person and their relationship. This database is relative to the database of oxygen respirator intelligent training check equipment used in colliery salvation. Unified management of personnel, equipment and test can be realized using bar code as the only sign. Abound information is conclude in the personnel files to convenient emergent condition as blood transfusion.

3 System Extending

At present, there are many problems with oxygen respirator intelligent training test equipment used in colliery salvation equipped with Chinese ambulance team. Sometimes, team members are not confident for the safety of the respirator on back and

unwilling to train under colliery. There is still no equipment which can check the dynamics performance indexes of the respirator completely. So, this intelligent training test equipment fills the domestic void. This system will change the backwardness of countries' safe rescue industry prompt the whole qualities guarantee the lives of the rescuers and reinforce the strength of country's rescue ability. This equipment is an innovation and considerable economy benefit to the company is wished.

This equipment has industrial test in some rescue teams, every item reaches the design request and have good effectiveness. The equipment offers way not only to prompt and explore the oxygen respirator intelligent training test equipment used in colliery but also the standard body quality parameters. A new discovers have been down here.

Since respirators has great use and in many fields, such as mountain, fire control, tunnel, petroleum, , chemical engineering and power safety, its training system has great use. Since it has gain national regard and great investment. Its extending is very great.

References

1. Dong, L., Yin, W., Ma, W., et al.: High-sensitivity, large dynamic range, auto-calibration methane optical sensor using a short confocal Fabry-Perot cavity. *Sensors and Actuators* 127(2), 350–357 (2007)
2. Sun, H., Xu, A.: Principle and application of MCS-51/96 serials SCM (Revision). Beijing Aviation University Publisher
3. Wang, R., et al.: Detailed explanations for Power Builder application and developed technology. Electron Industry Publisher, Beijing

The Intelligent Calibrator of Mine Gas Sensor

Jin-Yan Cao and Mu-Qin Tian

College of Electrical and Power Engineering, Taiyuan University of Technology,
West Yingze Street No.79, 030024 Taiyuan, Shanxi, China
tylgcjy@163.com, tianmuqin@tyut.edu.cn

Abstract. This paper presents a DSP-based intelligent calibrator of mine gas sensor, which is composed of micro-controller unit, signal conditioning unit, communication interface unit, keyboard and display unit, touch-control unit and gas circuit system. Based on CPU technology, electronic technology and control technology, this calibrator has overcome the problems existing in most of the current calibrators, i.e. low speed, low accuracy and labor-consuming, offering better services for mining safety.

Keywords: mine gas sensors, a calibrator, DSP, computer control.

1 Introduction

The gas sensors used in mine include methane sensor, carbon monoxide sensor, wind sensor, temperature sensor, and negative pressure sensor etc. As one of the most crucial parts, the sensor of gas is compulsory in the system of gas monitoring which plays an important role in the coal mining safety. The thermal catalytic component is widely applied to the current sensors for the checking of low-density gas at home and abroad. Since the component is a sensitive chemical device, zero drift and sensitivity changes will be common with the long hour working, affecting the reliability of the monitoring system. Thus improving the stability of catalytic sensor has been the focus of various studies. In order to improve the reliability of gas sensors, many designs of the devices has been offered and the components have also been selected from many sources. However, such problems will still exist in even the most sophisticated devices as device aging, temperature drift, electromagnetic interference, etc., leading to the unreliability of the system and posing a threat for the safety of normal and orderly production.

There are two main types of calibrators existing in the current coal mines: Remote-control and manual one. The latter is mainly composed of the cylinder of standard gas samples, hoses, flow meter and decompression valve. In the concrete checking, the sensor, flow meter and standard gas samples are all connected with hoses; and the valve is then manually adjusted so as to attain the standard gas flow rate of 200 ml/min. Finally, the measured value should be observed to make sure it is consistent with the value of the standard gas concentration, if not, a screwdriver should be used to

adjust the knob of the sensor so as to make the reading value of the sensor consistent with the that of the standard concentration. There exists the man-made error in the checking process by manually adjusting the standard gas flow. Furthermore, only one sensor can be checked at a time. Based on the manual calibrator, the remote calibrator is added the function of remote control, which makes the reading value of the sensor consistent with that of standard gas concentration through the remote control button. Therefore, the design of a smart calibrator of mine is an urgent problem to be solved. At present, the relevant domestic researches have been carried out and there are a number of products available. It can be convinced that the studies of the calibration of multi-sensors for the gases with different level of concentration will be the solution to the problems of mining sensors.

2 The Introduction of the System

With the combining use of such new technology as computer, microprocessor, automatic control, touch screen, digital electronics and signal processing, the intelligent multiplex calibration system can automatically or manually adjust and check 8 or more gas sensors, such as that of CH_4 , CO , CO_2 , H_2S , O_2 , negative pressure, wind speed at the same time. The intelligent multiplex calibrator is highly intelligent due to the realization of the functions of data collection, communications, query and analysis together.

The effects of man-made error can be reduced and the checking accuracy can be raised since the whole checking process is fully controlled by computers, such as the task of adjusting, data processing, judging of the results, except that sensors must be hoisted to the outside of line ball by the check operator. In addition, the flowing amount and channel of the air, CH_4 , CO and other standard gas can be automatically controlled. More importantly, it possesses many functions such as directing the input and output of the air, switching on and off the electricity, supporting the checking and adjusting of various alarm devices and sensors (such as that of CH_4 , CO , CO_2 , H_2S , O_2 , negative pressure, wind speed), including all kinds of mine sensors. It can make the changes in calibration process and corresponding relationships of data in line with different objects to be checked, conditions and the requirements of manufacturers. Still, due use of touch screen technology makes the operation more intuitive and convenient. Lastly, it can work on the automatic or manual checking of 8 or even more sensors.

Intelligent multiplex calibration station provides a DSP-based intelligent calibrator of mine gas sensor. It is possible to check and adjust the gas sensor since the gas sensor can transmit measurement data. According to national industry standards, in the process, the calibration instrument displays the value "zero" in fresh air or it can be adjusted to display the value "zero" by using gas sample. Under 1% ~ 2% CH_4 methane sample, the values displayed in the instrument should be consistent with that of the standard gas concentration, and the amount of flowing gas sample should be consistent with the requirements of product specifications. Therefore, comparing the measured data with the value of the standard gas concentration, the waiting time of the passing gas can be automatically controlled so as to make sure the measured data is the data stably displayed when the gas with standard concentration goes through the sensor. Meanwhile, comparing the value of gas flow with the given value, the decision to

control to constantly adjust the electromagnetic proportional control valve can be made in order to achieve stable and accurate flow control ($N \times 200\text{ml} / \text{min}$, where N is the number of sensors under test).

The technology and principles of intelligent multi-functional calibration consist of the principles of mine gas sensor, principles and features of the micro-controller, i.e. real-time controller (TMS320LF2407 DSP), the technology of computer platform (Power Builder), database (SQL Server) technology and control (prediction Control) technology, and the various parts work together to achieve the automatic checking and adjusting of multiplex of sensors of gases with different level of concentration. Fig.1 is the diagram of a micro-controller system of the calibrator. There are a wide range of mine gas sensors including the sensor of CH_4 , CO , CO_2 , H_2S , O_2 , negative pressure, wind speed. But as to the parameters, every kind of sensor generally involves measurement/input maximum, measurement/input minimum, transmission/ output maximum, transmission/ output minimum (transmission can be shown in frequency signal and the voltage signal). Different kinds of sensors are different in the range of input and output. For example, the input scope of the sensor of CH_4 input is within 10%, while that of CO can be ten or hundred times. To meet the requirements of different sensors, the project has designed a maintenance module, where the system

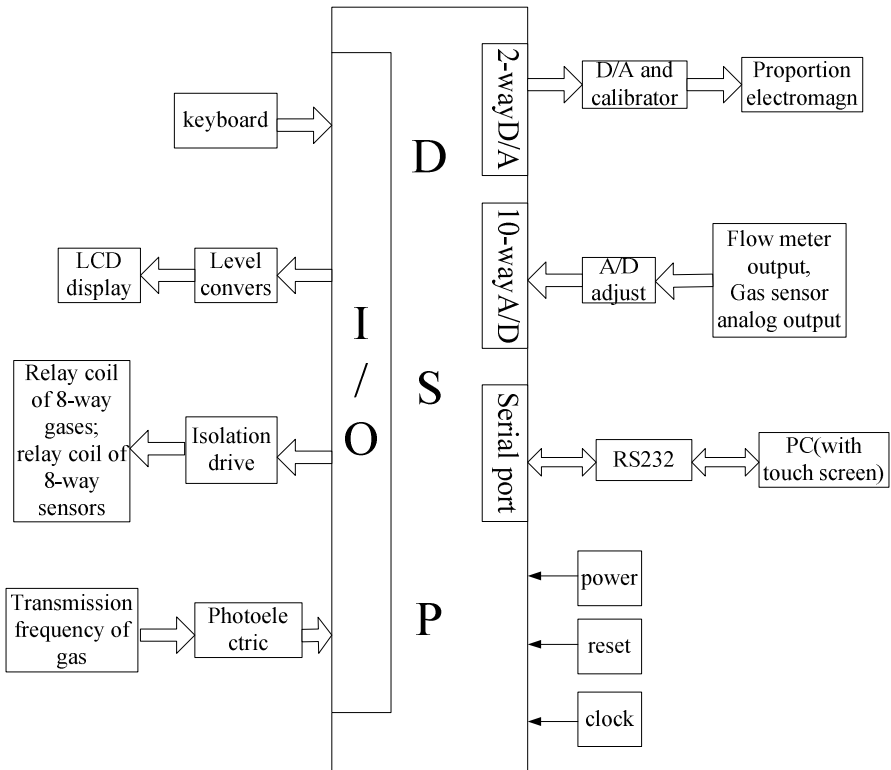


Fig. 1. The calibrator system micro-controller diagram

maintenance logs are created in a database. Based on the actual checking of the different sensors, the maintenance of them will be done so as to meet the calibration tasks of various types of sensors.

For the sensor with frequency as transmission signal, the microcontroller will collect the data through the P₁ port. Since the input current of P₁ port is low, it is isolated by Opt coupler. If the actual frequency transmitted by the sensor is f (Hz), micro-controller is to monitor the timing on the Edge. The time m will be transferred into the frequency: $f = 11.059 \times 1000000 / (12 \times m)$. And then according to the above measurement (input) maximum f_1 , measurement (input) limit f_2 , (transmission) output ceiling sc_1 , (transmission) output limit sc_2 , the actual value of concentration can be figured out: $mm = \text{abs}(((sc_1 - sc_2)/(f_1 - f_2)) \times (11.059 \times 1000000)/12/m - sc_1/4)\%$. Carrying out the same practice on every sensor respectively until the calibration 8 (even more) sensors are finished.

For the sensor with voltage as transmission signal, A/D conversion will be employed. And then according to the above measurement (input) maximum f_1 , measurement (input) limit f_2 , (transmission) output ceiling sc_1 , (transmission) output limit sc_2 , and the actual value of concentration can be figured out. With analog signals going through the RC low-pass filter, and a voltage accordingly adjusted by using difference proportion computation, on-chip ADC and DSPCPU are connected. According to national industry standards, if the sensors of gas with different level of concentration such as higher, mediate, and lower, are to be checked and adjusted, there are the requirements of the standard gas, the amount and the location of tested gas sensors. The microcontroller has 16-way switches, with full optical isolation, and the electric current of each way can be amplified to 500mA, which is enough to drive the work of relay. The output voltage can be selected among 5V/12V/24V based on isolated power supply, which controls the circuits of 16 relays (8-way of standard gas and 8-way sensors of gas with different concentration) to be switched on and off. According to the control signals given by the computer, the opening and closing of gas relays can be controlled. Likewise, according to the commands directed by the computer, the selected gas sensor is connected to the standard gas, which is then checked and adjusted under the operation of the automatic calibrator of the single gas or multi-gases or optionally intelligent calibration, followed by the task of the location of the sensor. In the actual practice, the position of the checked sensor is random, and the type of standard gas is finally selected. The computer will automatically record all the processes, and then send commands to the microcontroller.

After the data is received, the microcontroller will implement the order step by step: The first thing is to confirm the type of calibration. As to the calibration of multi-gases sensor, the number of the sensor should be firstly figured out, and then accordingly provides a stable standard gas (take CH₄ for example, CH₄ of low concentrations of, mediate concentration, high concentration) flow $N \times 200\text{ml}/\text{min}$ in order to achieve precise flow control. In the practice, each standard gas is pumped into the sensor in a certain time (about 90 seconds) to ensure that the sensor readings have been stabilized before getting the data of the sensor. At the same time, the remote control sensor calibrator is used to adjust the sensor so as to make the reading of it consistent with the standard reading. As to the automatic or manual calibrator of single gas, the commands will include the information of a sensor location, quantity and concentration of the gas. Based on the function, the location and quantity of the sensor to be checked, and the

concentration of the gas, the micro-controller will accordingly implement the tasks such as controlling solenoid valve to make the corresponding outlet way open, and making the path of the standard gas open, and controlling the voltage to the sensor under calibration so as to control the corresponding standard gas flow.

3 Conclusion

The application prospect of the product is very great. First of all, China is a big country of coal, and gas sensors are very popular. Secondly, gas checking is essential to coal mining. Thirdly, the components of gas sensor are easily subject to aging and drifting, the checking system will be likely to fail and the checking accuracy will be decreased, resulting in unnecessary losses without the due checking and adjusting. Fourthly, the gas sensor calibrators are highly appreciated by most mining enterprises since they are in small size, moderate in cost, easy to operator, and crucial to mining safety. Currently, all coal mines in China are gas mines, according to statistics, out of the 609 coal mines of the 100 key state-owned coal enterprises, the higher-gas mines account for 26.8%, the mines with gas outburst account for 17.6%, lower-gas mines account for 55.6%. Out of the local state-owned and township mines, higher-gas and gas-outburst mines account for about 25%. Apparently, Automatic calibrators will enjoy a very large market.

The essence of the calibration is the combination of checking and adjustment. Therefore, the future research should be focused on the further study of the principles and mechanism of the sensor itself so as to achieve the purpose of combining the task of checking, adjusting and analyzing together.

References

1. Lin, Y.-Y.: TMS320F240 × DSP assembler and C language function control applications, pp. 7–61. Beijing Aerospace University Press (2009)
2. Xiulin, W., Yunfeng, C.: Design of altimeter of micro air vehicle based on DSP. *Natural Science Edition* 35, 125–127 (2005)
3. Lei, D., Wangbao, Y., Weiguang, M., et al.: High-sensitivity, large dynamic range, auto-calibration methane optical sensor using a short confocal Fabry-Perot cavity. *Sensors and Actuators* 127(2), 350–357 (2007)

The Method of Power Saving and Study on Control System for Center Air-Condition Set

Guojun Zhang¹, Yao Zhao², and Chaofeng Liu³

¹ Information Engineering School, Taiyuan University of Technology, Taiyuan, 030024 China
zhangguojun606@sohu.com

² Automation School, Harbin University of Science and Technology, Harbin, 150080 China

³ Shanxi Taigang Stainless Steel Co., Ltd. Equipment&Material Department,
Taiyuan, 030008 China

Abstract. According to the function of recycled water in Center Air-condition Set and the run mode of recycle pumper, this paper proposes the method of power saving, discusses the corresponding structure of automation control system, draws up the control system block diagrams for the temperature control and pressure control, describes the system configuration for hardware, as well as provides the method in programming for the temperature data sampling, digital filter, digital controller and etc.

Keywords: Recycle Pumper, Power Saving, Control System Block Diagram, Temperature Data Sampling, Digital Filter. Temperature unit.

1 The Function of Circulating Water and the Run Mode of the Pumper

When Center Air-condition works, with the work mode of circulating water, not only the freezing water refrigerated by air conditioning units in summer, but also the heating water heated by air conditioning units in winter, is sent to the user area by the water circulating pumper of the units. Circulating water is sent out through the user area, then back to the air conditioning units, in which to be refrigerated (in summer) or be heated (in winter), and sent out again and again. So, when Center Air-condition works, it is must to start the Circulating Water Pumper first. Only under the premise of the Circulating Water Pumper running, may other related equipment in the units start to run, and when the Circulating Water Pumper begins to work, it still run until the end of the season

2 Power Saving Design for the Running Mode of the Circulating Water Pumper

In the whole year, no matter summer or winter, the change of weather is regular. For example, in summer, maybe the sun is shining with high temperature yesterday, but it

rains with cool weather today. Regardless of summer or winter, the change of surroundings temperature is great during the working period of the whole air conditioning, which from the start of summer (winter) to the hot weather (the cold weather), then till the end of summer (winter). So for the purpose of power saving, it can choose a Circulating Water Pumper to do real-time adjustment and operation according to the change of weather. For example, when the temperature is relatively low in summer, the air condition usage in the user room may be relatively reduced. At this time, it can reduce the running speed of the Circulating Water Pumper and the throughput of the cool air to reduce the load when the units refrigerate, as well as to save power when the Circulating Water Pumper runs. Similarly, when the weather is relatively warm in winter, the air condition usage in the user room may be relatively reduced. At this time, under the premise of reducing the heating load of boiler, it can reduce the running speed of the Circulating Water Pumper, the throughput of the heating water and the power consumption of the Circulating Water Pumper to save power. According to the characteristics of the pumper (equation 1), once the speed of the pumper reduces, the relation of the equipment energy consumption P and the energy consumption which not used in speed regulating P_e is cubic relationship of prompt drop rate, so once the speed is reduced to 50% of the rating value, the power consumption is reduced about 49%.

$$P = (n/n_e)^3 P_e \quad (1)$$

Usually the selection of the Circulating Water Pumper is designed according to the maximum load. For example, according to the summer and winter in Shanxi, China, the temperature difference in a day is about 10-20°C, so it is possible to save 30%-40% power consumption of the pumper by speed governing about the Circulating Water Pumper. The conclusion can be explained in most area in north of Chinese Mainland. As to the south of Chinese Mainland and the coastal areas in north, although the temperature difference in summer is not the same in other area of north, this run mode of saving power is very suitable during the start period and the later period of air conditioning works, as well as in the condition of the temperature mutations.

3 Power Saving Control System Structure of the Circulating Water

3.1 The Determination of System Design Scheme

Firstly, in order to realize computer control, the system chooses PLC as main control computer to realize digital control. Secondly, the speed regulation of the Circulating Water Pumper is realized by the AC-DC-AC inverter. Thirdly, besides making the Water temperature sampling for the effluent water and the backwater of the Circulating Water as the temperature control feedback signal, it should detect the ambient temperature to design the reasonable control target value of the circulating water temperature.

3.2 The Data Acquisition and Processing of Temperature Signal

The special temperature modules supporting programmable logic controller, such as PT100 or thermocouple, were selected in system design for temperature data conversion. Its advantage is that thermal resistance or thermocouple temperature signal can be directly read to the specified storage address or the corresponding cache register of CPU units, eliminating temperature transmitter link. Thus, on one hand, we can save money, on the other hand, the system structure is more simple and reasonable. When monitoring the data, if the binary code reading is inconvenient,, it can be converted into BCD yards according to the related instructions and then take other data processing [1]. Considering electromagnetic interference and other environmental factors in the master driver running, the corresponding digital filtering link should be adopted after the temperature data sampling in programming. The arithmetic average weighted filter[2] filter method not only can solve signal interference problems but also make the filter programming to read easily in filtering. The first-order lagging filter have not been selected [2] for filter in system design, because when the master driver running, the PWM pulse width modulation frequency often change, that will make filtering bandwidth calculating complex and digital filter not easy to attain and programming trivial. In addition, the data after sampling is the binary code, so the system should be increased scale transform in the feedback channel. Thus the binary code detect data will transform the dimensionless feedback temperature data [2].

3.3 The Block Diagram of Temperature Control System

According to the design of system plan, pumper frequency startup’s control characteristics, automatic control principle theory [3], we can give figure 1 which shows the central air conditioning circulating water temperature automatic control system. The given temperature value in block diagram is not a direct temperature setting, but a adjust result according to the environment temperature data, a backwater temperature measurement value, practical experience. The control temperature feedback signal, considering the outlet water temperature and backwater temperature of circulating water, is also calculated.

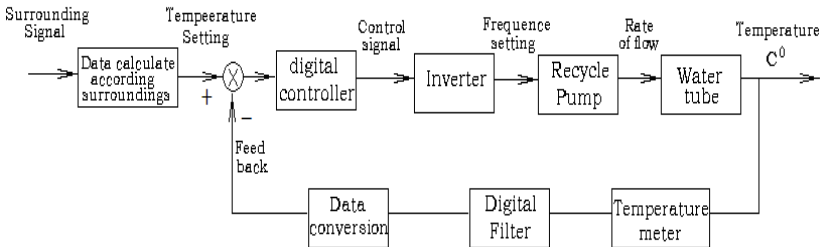


Fig. 1. Single Loop Control System Block Diagram

3.4 The Double Closed Loop Control Structure in This System

Without considering the pipeline pressure conditions, the circulating water temperature control process is usually similar to the constant pressure water supply system, the PI or PID regulation will meet control requirements when controlling [4]. Adopting this controlling, when the system needs to reduce circulation water flow so that the pumper rotation speed reduce slowly, its shortcoming is that the circulating water pipe stress will reduce a lot, This will bring two kinds adverse consequences for equipment and running, one is that it will cause equipment failure and even damage, because the backwater of circulating water is heated by heat exchanger in winter, if the water velocity low and without temperature automatic regulation control in heat exchangers, the heat exchanger will overheating; The second is, when lowering rotate speed, than it may make the pipeline network pressure insufficient so as produce gasification phenomenon. Above two phenomena are not allowed to happen. A method to solve the problem is adding a pressure control inner loop (figure 2), according to the double closed loop control method in electric drive control system [5], based on the structure of control system shown in figure 1, or with velocity instead pressure control loop. Since the dynamic process of the corresponding slower than DC motor drag in this system controlling, and load changes slower, current wave little, this add a current control loop in master driver, which similar to current control loop in DC motor drag control system, and the system becomes a three-ring system, so it is very tedious and unnecessary.

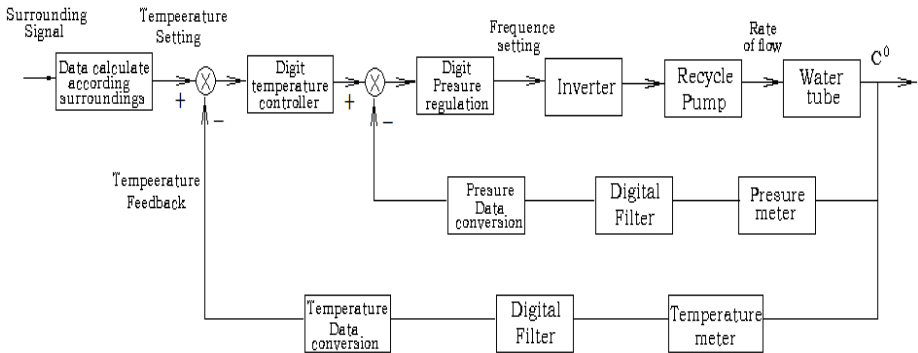


Fig. 2. Double Loop Control System Block Diagram

The digital controller in control system block diagram adopts PI or PID adjustment algorithms of the most pumper control system. In the specific programming, considering PID control parameters adjustment convenience in equipment operation period, the PID instructions of PLC instructions is no used, while PID algorithm formulas are programmed according to the PID instructions, the corresponding PID control procedures.

4 The Hardware Configuration of the System

First, in order to manipulate and manage conveniently, a touch screen is chosen which can be used to show temperature data, control equipment start or stop, instruct running conditions, do manual intervention control and adjust temperature's setting[6]. In addition, we can choose Mitsubishi FX series or Siemens S7-200 models as PLC. This article use Mitsubishi as an example. when the system has two circulating pumpers, one of which is on use and the other is spare, the host can choose FX1N - 24M, in consideration of the system has about eight to ten input points that are two pumps' operation contactor contact signal of frequency and variable-frequency, variable-frequency inverter' signals of running and fault, switch signal as input signals of program control switch and about six to eight output signals that are four pumps' contactor driver control, frequency converter's running and fault reset and so on. If the system has three or four pumps, the host should be chosen FX1N-40M or FX1N-60M (figure 3); the temperature module should be chosen FX2N-4AD-PT of PT100 RTD module which can be input four PT100 temperature signals; In addition the design uses a FX0N-3A module, which has two input channels and one DA channel. AD can be used respectively for pressure monitoring and output frequency detection of the inverter. DA can be used for inverter's speed control. If the variable-frequency inverter has pulse output signal which can be used to represent the current frequency output value, the pulse output signal can be sent to X0-X7 that are high-speed counter inputs of the FX1N host. By High-speed counter we can also realize the detection of the value of the inverter output frequency. The design of digital (binary) input and output can refer to the general PLC logic control approach.

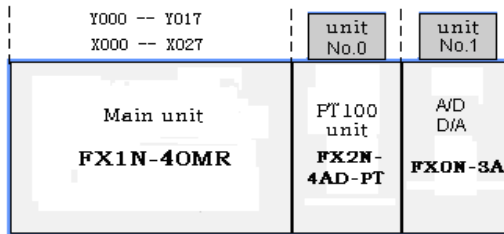


Fig. 3. Construction of Controller

5 Conclusion

According to the above-mentioned method, design an energy-saving control system of central air conditioning for circulating water, which was implemented in the Royal Palace Hotel in Taiyuan. The Circulating pumper power of its high area is 132KW and the circulating pump power of its lower zone is 110KW. When the equipment was put into operation, it is coincided with the coldest season in North China in January. The examined results is that the average daily cost saving is up to more than one thousand Yuan. When the similar systems in Shanxi Daily editorial building were put into operation, it also received a good energy-saving effect of automatic control's quality and products.

References

1. Keming, X., Luyi, X.: PLC Theory and Program Design., vol. 6, p. 235. Publishing House of Electronics Industry Press (2010)
2. Xinmin, P., Yanfang, W.: Microcomputer Control Technique. Higher Education Press vol. (10), pp. 213,214,218 (2000)
3. Long, Y., Guojun, Z.: Adaptive Variable Structure Control System, vol. 10, pp. 9–10. Development and application of computer Press (2004)
4. Guojun, Z.: WF-III Intelligent Digital Multi-Parameter Control Frequency Control Automatic Water Supply Device, vol. (1), pp. 140–142. Industrial Control System Application Press (1996)
5. Boshi, C.: Electric Drive Automatic Control System, vol. (1), pp. 51–52. Machinery Industry Press (2010)
6. Jinlong, Z., Guojun, Z.: Human-Machine Interface Design of Center Air Conditioning Circulating Water System. Mechanical Management and Development (1), 98–99 (2011)

Design of a Temperature and Humidity Monitoring/Sending System Based on SHT1 Temperature and Humidity Sensor

Jun Guo, Mu-Qin Tian, and Mu-Ling Tian

College of Electrical and Power Engineering, Taiyuan University of Technology,
West Yingze Street No.79, 030024 Taiyuan, Shanxi, China
tianmuqin@tyut.edu.cn

Abstract. This paper introduces a temperature and humidity monitoring/sending system, which is based on digital technology and computer technology, using the SHT1x sensor made by Sensirion company of Swiss. The system overcomes disadvantages of poor linearity, low accuracy transmission complex use of traditional system with the analog humidity sensor and makes measurement for temperature and humidity higher precision.

Keywords: SHT1x sensor, temperature and humidity, monitoring.

1 Introduction

Traditional analog humidity sensor need signal circuit design, adjustment and calibration, for which, the precision can not be guaranteed, so does the linearity, repetition, interchanging and consistency. SHT1x is a new type temperature/humidity sensor made by Sensirion Company based on CMOSens technology, which combines CMOS chip with sensor. Temperature /humidity sensor, signal magnifier, A/D switch and 12C bus interface are intergared in one chip. The author designs a temperature/humidity measurement and sends system with SHT1x sensor for temperature/humidity measurement in close circumstances. Live temperature or humidity value is sent to a forane receiver through wireless signal.

2 Hardware Circus Design

ATS89S52 SCM was used as CPU in this system, the frame is shown in Fig.1. SHT10 was choosed as temperature /humidity sensor. SHT11and SHT15 also products belonging to this series. Technology indexes are shown following [1]:

Table 1. The technology indexes

Type	Humidity precision (%RH)	Temperature precision (°C)	Encuplation
SHT10	±4.5	±0.5	SMD
SHT11	±3	±0.4	SMD
SHT15	±2	±0.3	SMD

Since AT89S52 has no I2C serial interface bus, P3.6 and P3.7 are used separately to simulate the timepiece SCL and date line SDA of I2C. AT89S52 is used as host computer, SHT10 is used as secondary computer. Communication is realized through the simulation I2C bus interface. SCM uses AT89S52, which is the upgrading AT89C52, adding some function such as ISP downloading, WDT, and so on. Time of temperature /humidity measurement is supplied by clock chip, measurement parameters can be stored in 31 RAM units.

Display uses RT1602C LCD, two lines shown, with 16 characters per line. Each character size is 5×8t lattice. Temperature, humidity, time and data are shown.

And the humidity value are sent to the data port to code. An integrated code composes of address code, data code and synchronization code is sent to the emitting module which has an emitting frequency of 315 MHz and distance of 70 m. AM-RT4-315 is used here.

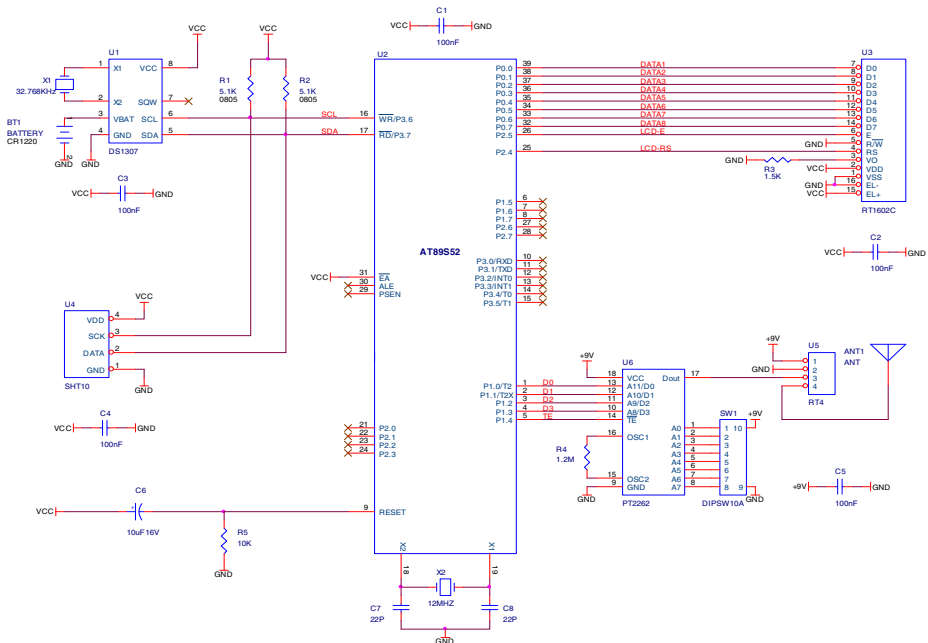


Fig. 1. The system hardware principles diagram

3 Temperature and Humidity Sending Wirelessly

TXD of PT2262 is only 6 Bit(D0~D5), can not be sent as Byte, but as Bit, meaning “0” or “1”. Temperature and humidity data from this system is 4Byte (2 Byte temperature and 2 Byte humidity). If they are sent as Bit, total 32 Bits not only have long sending time, but also face data losing and disturbing. Synthetically thinking, D0~D3 bits send 4Byte through 1/2 Byte type.

Subprogram for temperature is shown following:

```

SEND_TEMPERATURE:
MOV  A,TEMP_H
ANL  A,#0F0H
SWAP A
MOV  P1,A
CLR      TE; dicimal of temperature
LCALLDLY150MS
SETB TE
LCALLDLY20MS
MOV  A,TEMP_H
ANL  A,#0FH
MOV  P1,A
CLR      TE; figure of temperature
LCALLDLY150MS
SETB TE
LCALLDLY20MS
MOV  A,TEMP_L
ANL  A,#0F0H
SWAP A
MOV  P1,A
CLR      TE; decimal fraction tenth of temperature
LCALLDLY150MS
SETB TE
LCALLDLY20MS
MOV  A,TEMP_L
ANL  A,#0FH
MOV  P1,A
CLR      TE; decimal fraction hundredths of temperature
LCALLDLY150MS
SETB TE
LCALLDLY20MS
RET

```

4 Software Program

One important part of this software program system is temperature and humidity sampling and nonlinear compensating of SHT10. The default measure resolution of SHT 10 is 14 bit(temperature) and 12 bit(humidity). This can also be done through state register lowering to 12bit and 8bit. Following are nonlinear compensating functions separately for temperature and relative humidity.

1) Relative humidity nonlinear compensating computing:

$$RH_{linear}=c1+c2 \cdot SORH+c3 \cdot SORH^2 \quad (1)$$

Where, SORH is the humidity value of sensor, and coefficients are shown here:

12 bit SORH: $c1=-4$ $c2=0.0405$ $c3=-2.8 \cdot 10^{-6}$

8 bit SORH: $c1=-4$ $c2=0.648$ $c3=-7.2 \cdot 10^{-4}$

2) Temperature computing:

SHT10 temperature sensor has good linearity, so the following functions can change the temperature value from digital number to actual value:

$$\text{temperature}=\text{d1}+\text{d2}*\text{SOT} \quad (2)$$

Where, SOT is the temperature value of sensor, when voltage is 5V, resolution is 14 bit and 12 bit, the coefficients are shown here:

14 bit SOT: $\text{d1}=-40$, $\text{d2}=0.018$

12 bit SOT: $\text{d1}=-40$, $\text{d2}=0.072$

Sampling software is written by assemble language, which has a character of simple interface operation, but complex compute of relative humidity. Manufacture only offers sampling subprogram of C language. The author integers the function (1) and writes temperature/humidity sampling and nonlinear compensating subprogram.

5 Conclusion

Measurement system of this paper is fixed in a closed environment with high temperature and high humidity, temperature/humidity value is sent 5 second per time. Temperature/humidity value is shown livingly through a set of receiver display.

References

1. Sensirion Inc. SHT1x DATA SHEET (2007)
2. ATMEL Corp. AT89S52 DATA SHEET (2008)
3. Slotions Ltd. AM-RT4-315 DATA SHEET (2008)
4. Princeton Technology Corp. PT2262 DATA SHEET R.F

Research and Implementation for a Content-Based Video Playback System

Zhang Lin, Duan Fu, and Li Gang*

Taiyuan University of Technolog, College of computer and Technology,
Taiyuan Shanxi 030024 China
Ligang@tyut.edu.cn

Abstract. Monitor video playback should get specific scenario by way of either traditional drag-and-drop or location on time line for long time, which is hard to search for necessary information fast and accurately due to mass of video data. However, the method used present in the paper that employs, color histogram to detect short-cut and implements content-based video playback by analyzing key frame image.

Keywords: Color Histogram; shot-cut detection; Content-Based Video playback.

1 Introduction

It is an information age in the 21st century and multi-media information has become a major source of date increasingly. Except to traditional text information, a large number of sound, image and even video data are widely used by people. How to manage and use these dates effectively has become an urgent problem to solve. Content-based visual Retrieval (CBVR), obtained intensive attention and being researched inclusively by scientist or technician home and abroad in the filed of multi-media information, has come about along with the new information century and quickly become a hot spot in the research of multi-media, especially in the research of image, video and database technology.

CBVR technology can be widely used in various industrial, scientific fields, such as remote monitoring, multimedia conferencing, virtual reality, TV news, satellite images, information entertainment, interactive digital television, interactive e-commerce and video digital library, what worth mentioning especially is interactive digital television and video digital library, their broad developing prospect and huge commercial value is the powerful booster of its vigorous development.

This paper has used color histogram to implement Content-Based Video playback system VideoRetrievalPlatForm.

* This work is supported by the Fundamental Project torch-plan projects of Shanxi Province No.20080322008.

2 Color Histogram

2.1 Color Spaces

Colored and no colored images are two types of image color [1]. No colored images are being replaced increasingly by colored images due to the advances of technology. According to the human eye structure, all colors can be seen as different combination of the three basic colors - red (R, red), green (G, green) and blue (B, blue). Researchers have proposed many kinds of color spaces [2] one after another, such as RGB space mainly used for color display or HSI space mainly used for color processing, etc. most color image displayers has used RGB images, more image formats of windows operating system have also used RGB model as stored image datas. Normalized RGB model, because it can partially avoid the influence of shadow and illumination changes on the colors, is widely used in the extraction of video image feature, the definition is as follows:

$$R_N=R/S, G_N=G/S, B_N=B/S, S=R+G+B \quad (1)$$

2.2 Histogram

Histogram describing color distribution of an image, which is the most reliable feature of the image. The core idea of Color Histogram Method [4] [5] is to statistic the frequencies of any color on an image appearing in a certain color space. Firstly, dividing the color range into n discrete interval, calculating the number of pixels per frame in an image fell into each interval, and obtaining its color histograms, than calculating the histogram difference between two frames in an image as a distance measure.

Global color histogram matching algorithm uses the statistics of an entire image pixel to calculate the difference between the image frames. If we divide color space into n interval, $h_i(k)$ and $h_j(k)$ separately represent pixel numbers that the i-th and the jth frame fell into the kth color interval, so distance between frames measurement function can be defined as follows:

$$D(x_i, x_j) = \sum_{k=1}^N \left\{ \left| h_i(k)^R - h_j(k)^R \right| + \left| h_i(k)^G - h_j(k)^G \right| + \left| h_i(k)^B - h_j(k)^B \right| \right\} \quad (2)$$

This system uses global color histogram of NRGB Space. According to the research of bibliography [6], the division of sub-intervals and determination of shot-cut conversion detection threshold T are as follows, histogram divides 0-255 color space into 64 sub-intervals, and select 5 times all frames' average difference ADF as shot-cut conversion detection threshold, which can guarantee a higher recall and precision.

3 System Design

3.1 Functions and Modules

VideoRetrievalPlatForm, using the technology of color histogram and relying on DirectShow SDK to analyze data, has achieved many functions such as video play, video

location forward or backward on frames, shot-cut detection, key frames storage ,video playback according to key frames, etc. Basic modules have been shown in figure 1:

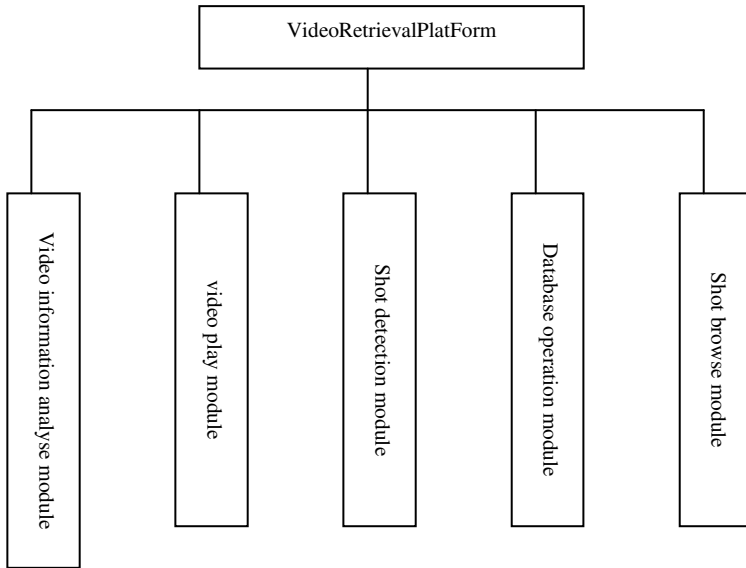


Fig. 1. System modules

3.2 Brief Introduction of DirectShow

As the system involves in the operation and analysis of original data stream of video media, So we chose Microsoft's VC ++. Net 2005 platform, as we all know, VC ++ is powerful for image and multimedia process. We use Direct Show SDK (a member of DirectX 9.0 SDK) to acquire and analyze video frames.

DirectX SDK which is provided by Microsoft and developed upon Windows operating system is a set of interface of the high-performance of graphics, sounds, input, output and network game programming. DirectX is defined as "device-independence" by Microsoft, that is, the use of DirectX can provide a high performance of device-dependence by device-independent way.

The design processes of the system mainly bases on the process framework to video by Direct Show, A traditional DirectShow system is shown as fig.3, DirectShow located in the application layer of the system uses a model called the Filter Graph to manage the entire data stream processing; each function module involved in the data processing is called Filter; each Filter in the Filter Graph has been connected into a "pipeline" according to certain order to work together.

Filters can be divided into three types by function roughly: Source Filters. Transform Filters and Rendering Filters. Source Filters are mainly responsible for getting data, the data source can be a file, the capture card in the internet computers, digital cameras derived by WMD or VFW, then transfer the data downwards; Transform Filters are mainly responsible for data format conversion, Such as data separation / synthesis, decoding / encoding, etc, then transfer the data downwards continually; Rendering Filters are mainly responsible for the final destination—send

data to the graphics card and sound card for multimedia presentation, then transfer the data to a file for storage. A typical DirectShow framework is shown in Figure 2.

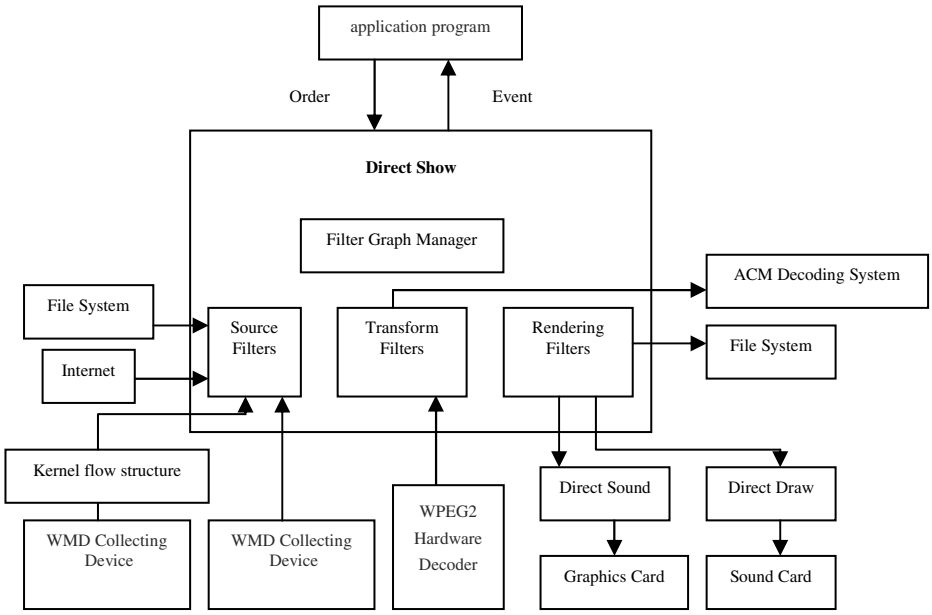


Fig. 2. DirectShow framework

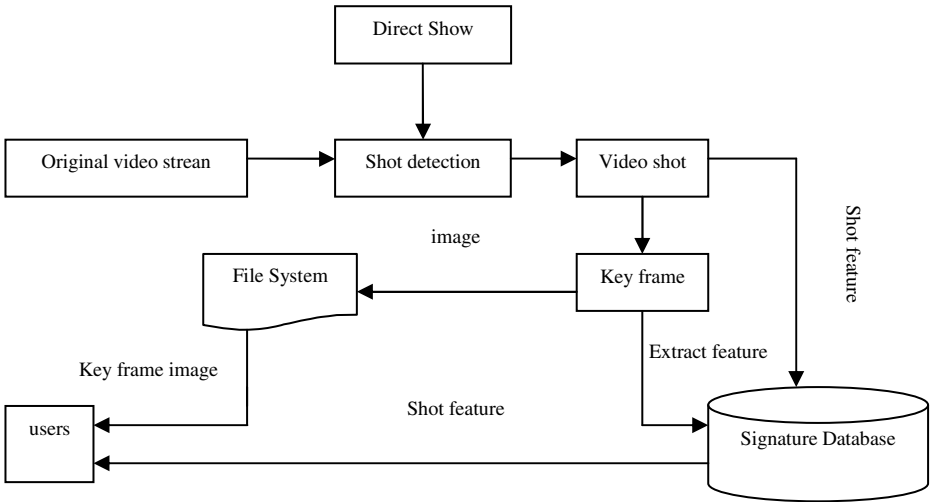


Fig. 3. Content-Based VideoRetrievalPlatform framework

3.3 System Framework

The design of VideoRetrievalPlatForm abides by the general model of video shot retrieval, shot-cut detection algorithm bases on global color histogram, and data analysis depend on DirectShow framework.

3.4 The Design of Class in the System

In order to make the system has good expansibility and manageability; we use the design idea of class. According to the operation principle of Direct Show and the need of system analysis, we have designed ten classes mainly in the experiment, shown as table 1:

Table 1. Class design of VideoRetrievalPlatForm

Class name	Function of the class
CPlayStream	Analysis of video information, and control the position when video is played
CMedialHandle	Interception of video image
CImageShow	Show of key frame image
CDataBaseControl	Database operation
CVideoShotMatch	Shot matching
CVideoPartitionClassOneDlg	Direct pixel Differencemethod
CVideoPartitionClassTwoDlg	Shot partition algorithm of global color histogram
CVideoBrowseDlg	Content-Based shot browsing
CErrorCodeShow	Error information processing

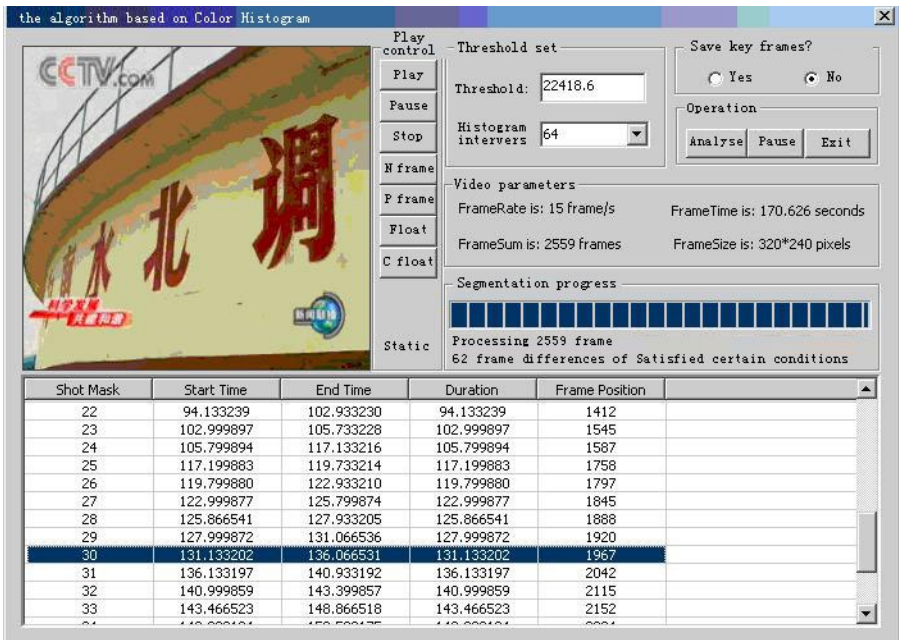


Fig. 4. Video shot-cut detection

3.5 System Interface

Shot-cut detection needs to firstly set up video path which shall be processed and analyzed, then operate global color histogram based shot partition algorithm to detect shot-cut, and store key frames of the shot at the same time, as shown in figure 4:

After partitioning the shot automatically, we can browse the video in accordance with the shot. Turn on a news video named xinwenlianbo_1X.wmv, and below the video there will automatically list all its 91 key frames' images been detected, Click on the image, browsing a video quickly can be achieved. Fig.5 is clicking on a key frame named 66.599930_68.333265.bmp to achieve the video play.



Fig. 5. Use key frame to achieve the video play

4 Conclusion

This paper has researched content-based video Retrieval and implemented a system called VideoRetrievalPlatform by a series of technique such as VC++.NET, COM,

DirectShow, etc. The system has many functions such as extraction and analysis of basic information of the video, video play, video location forward or backward on frames, shot-cut detection, key frames storage, video playback according to key frames, etc. The results of this paper have practical application value for monitor system especially. What we often interested in are those cutted shot in the video monitor system, video playback according to key frame which can not only quickly implement the location of specific shot, but also help workers analyze the emergencies.

References

- [1] Shuming, Z.: Image-image processing and analysis of engineering, pp. 17–20. Tsinghua University Press, Beijing (1999)
- [2] Gevers, T., Smeulders, A.W.M.: Color-Based Object Recognition. *Pattern Recognition* 32, 453–464 (1999)
- [3] Qingming, H., Tianwen, Z., Shaojing, P.: Color Image Segmentation Based on Color Learning. *Journal of Computer Research and Development* 32(9), 60–64 (1995)
- [4] Schulte, S., De Witte, V.: Histogram-based fuzzy colour filter for image restoration. *Image and Vision Computing* 25(11), 1377–1390 (2007)
- [5] Fang, H., Jiang, J.: A fuzzy logic approach for detection of video shot boundaries. *Pattern Recognition* 39(9), 2092–2100 (2006)
- [6] Li, G., et al.: Research on Threshold of Shot Segmentation based on Color Histogram. *ComputerDevelopment &Applications* 16(11), 9–10 (2008)

Research on Updating Land Use Database Dynamically by Remote Sensing Monitoring

Xiuming Jia and Qiang Li

College of Mining Technology, Taiyuan University of Technolog,
030024 Taiyuan, China
xiumingjia@163.com, sxliq2119@126.com

Abstract. The variation of land use monitored by dynamically remote sensing can be calculated by gray value of corresponding inter-temporal image pixel difference. The image subtraction method can enhance the remote sensing images, which can highlight the variation data, and then dynamically monitor the variation. This Paper selects Xinrong District, Datong as an example, by using image interpolation method to deal with two remote sensing images for interpolation analysis, and then extract land change information. After the verification of field measurement, removing pseudo changing pixel, the land use database is updated, which achieve the purpose of using remote sensing method to monitor land use database dynamically. The experimental results show that the image interpolation method for extracting the variation of inter-temporal remote sensing images of land use is fast, accurate and practical.

Keywords: dynamic monitoring, RS, image interpolation, land use database.

1 Introduction

The land on this planet of human survival is the non-renewable resources, and is the basic elements of a national founding [1]. With the rapid development of China's economy, land resources that can be taken advantage will be less, so the conflicts between economic and land using will be become increasingly prominent. In order to be better planning and land using, monito the land use real-timely, maintain the land-use database timelinessly, we must regularly update the database and change. The remote sensing technology were used to monitor land use change because of it's widen observation range, quickly accessed the data and less human disturbance, to provide timely and effective changes in the data to update the base year land use database.

This paper processed gray D-value operation about multi-date images, combined with land use status maps extracting variation information, and completing full variation information by outside survey. Then update the land use status database and provide all of the monitoring figure spots in written and analysis report of changing.

2 Study Area and Data Source Selection

2.1 Study Area

The XinRong District of Datong City is located in the north of Datong City, Shanxi Province, under the jurisdiction of a town and six townships (XinRong Town, Guojia Yao Township and Baozi Wan Township, PoLu Township, Shangshen Jian Township, Xicun Township, Huayuan Tun Township), including 142 administrative villages and 167 villages. It locates in the north latitude $40^{\circ}07'30''$ - $40^{\circ}25'00''$ and longitude $112^{\circ}52'30''$ - $113^{\circ}33'45''$, shown in Fig.1.

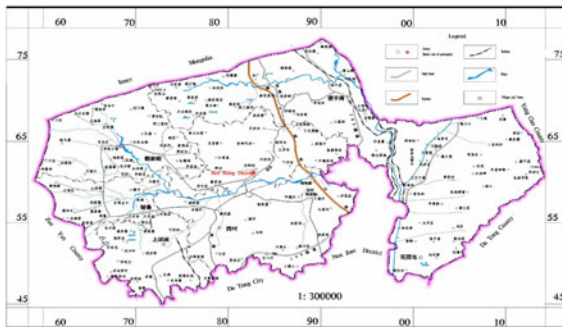


Fig. 1. Map of Xinrong District, Datong City

2.2 Data Source Selection

The data used in this experiments include: the remote sensing data and the ancillary data .The remote sensing data include: the second national land survey base maps marched by the China Aero Geophysical Survey and Remote Sensing Center Land Resources in 2008 which is 1:10000 production standard framing (The Image is getting in October 2007, the resolution of 0.5 meters, and color infrared aerial), the map image which is for the second land survey and getting in October 2009.The supporting information include: the administrative boundaries and the land use database(2008).

The image data analysis: the texture of the two chosen remote sensing images are clear, the side view angle is 10° ,the cloud amount is 2%, the key urban-rural area has no clouds, which are correspondent to the indicators in the regulations for remote sensing monitoring land use(the side angle is less than 27° ,the cloud amount is below 10%[2]).The phases of the two remote sensing images are basically the same, based on which, land use changing information can be easily extracted. The vector data of land use is completed and the drawings are clear, so, these can be used as the standard base maps.

3 Dynamic Monitoring Technology Roadmap

Dynamic monitoring of land use is the second national land survey in the two remote sensing images as the main data source, supplemented by land use database and

administrative boundaries and other auxiliary data. Corrected by remote sensing images, mosaics, fusion and other processing steps, it can produce the high spatial resolution and the high spectral resolution remote sensing orthophotos. Automatically processed by computer and manual interpretation of a combination of methods, compared to two phases of the remote sensing orthophoto, the land use change information can be extracted. Through the field investigation, the extracted information which obtains in the industry can be verified, and the missing polygons within the industry can be added. After the post-processing within the industry, the land use database can be updated. Xinrong district land use change from 2008 to 2009 in Shanxi province is an example for dynamically monitoring land use in the study area, the dynamic monitoring process shown in Fig.2.

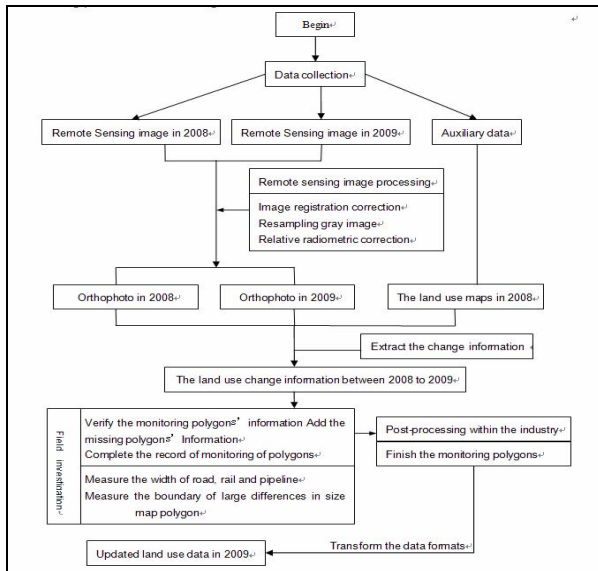


Fig. 2. Dynamic monitoring process

3.1 Remote Sensing Image Processing

As data is collected through the radiometric correction of remote sensing products, the pretreatment of remote sensing data in this study include the image registration correction, image resampling and gray-scale image enhancement.

3.1.1 Image Registration Correction

When the images in the geometric position changes, resulting in the ranks such as uneven ground pixel size and the size of the corresponding inaccurate, irregular changes in surface features distortion, that shows geometric distortion occurred in remote sensing image [3]. There are a variety of geometric distortion correction methods, but the most commonly used is a common method called precision calibration.

Geometric correction [4] is the use of ground control points (GCP) due to random factors from a variety of geometric distortion correction of remote sensing images. GCPs should be selected based on the object to registration and the number of GCP should be more than $(n + 1)(n + 2) / 2$ (n is the number of polynomials). The study selects nine GCPs, that the number meets the requirements, and uses the ARCGIS9.x software to regulate the images. Then the max control point residuals are 1.05 and the total RMS is 0.84.

3.1.2 Resampling Gray Image

To determine each pixel brightness value of the corrected image, it needs for resampling gray image. Currently there are nearest neighbor method, bilinear interpolation and cubic convolution.

Considering the calculation for bilinear interpolation method is simple, the sampling precision and geometric correction is in high accuracy, and the images brightness is continuous after the correction, while the spectral information of the image can be changed, which will likely to cause the loss of high frequency components that led to make the images blur, but for the whole study, bilinear interpolation for grayscale resampling image should be adopted.

3.1.3 Relative Radiometric Correction

Relative radiometric correction can be reduced the impact which caused by the differences between multi-temporal remote sensing images such as the atmosphere, illumination, phenology and sensor calibration, helping improve the dynamic monitoring accuracy. Generally speaking, a phase image accomplishes a reference image that is the main image [5], and another phase radiometric calibration image is the from image [5], that makes up the main - from the image pairs. Use of the image pixel gray value, the establishment of multi-temporal remote sensing image calibration equation between each band, normalized to the remote sensing image processing.

With this experiment, the final choice is the remote sensing images in 2008 as the base year data to determine change information, so the main - from the image pairs selects the image in 2008 as the main image and the image data in 2009 as the from image.

3.2 Extract the Change Information

3.2.1 Extract the Changed Pixel Information

Histogram matching. Match the 2008 single-band image's histogram with the 2009 single-band image, making the brightness distribution of the two as close as possible.

Difference operation. Using gray value difference between the corresponding pixels of two different time phase remote sensing images after registration correction, radiation correction and gray re-sampling, highlight the changed part by computer automatic processing methods, generating differential value image.

Vector of change information. Vectorizing the area that have changed obviously in the different value image, extract the variation information combined with land utilization present situation data, generating variation of information spot. The remote sensing image is as follows.

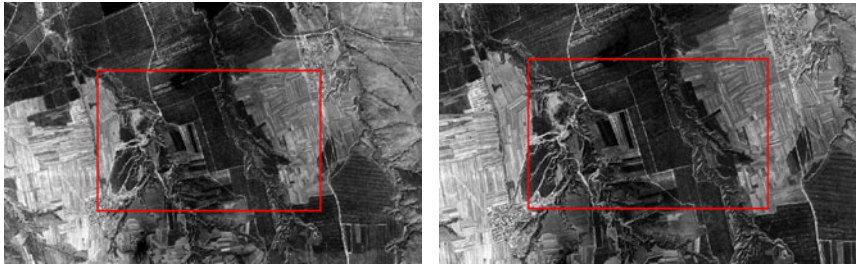


Fig. 3. Orthophoto map in 2008 (The left image's histogram does not match, and the right has done)



Fig. 4. The left image is Orthophoto map in 2009, and the right image is the Difference Image

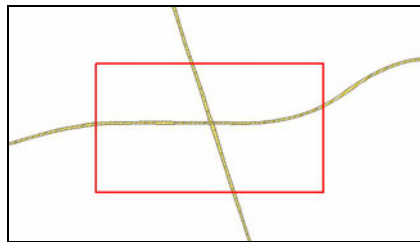


Fig. 5. Changed pixel between 2008 and 2009

Base on this, combined with 2008 land use maps, the changed polygons must be extracted and got its serial number one by one. The minimum area of the changed polygon must be correspondent to the rule of 《Technical requirements of land use change survey》 made by the ministry of land and resource. The area of population center is 4mm^2 ; arable land, garden plot and swag are 6mm^2 respectively; woodland, grassland and unexploited land are 15mm^2 respectively.

Overlay the vector data in 2008 land use database with satellite image, and then annotate the class code of vector data; after that do the visual interpretation of images and vector data, and verify whether the two data is the same or not. If the data is not the same and the changed polygon areas is more than the minimum area, and then collect the ranging line of the polygon according to the satellite image. Compared with the satellite image, the obvious ranging line of annotation and rendition should

be in the range of $\pm 0.2\text{mm}$, and the unobvious ranging line should be $\pm 0.5\text{mm}$ [6]. If the changed polygon has two different land types, the land type ranging line can be used to break up the polygon.

3.2.2 Statistics of the Changed Polygons

After the collection of changed polygon, attribution assignment should be done on changed polygon. The attribution should conclude location, ownership units, code number [7], polygon number, farmland slope rating, land rage coefficient. And in the land updating survey, the polygon unit should be a village or village people, besides only the nature of ownership, classification number, polygon number, land slope should be coding, shown in Table 1.

Table 1. Record of Each Patch

Polygon No.	Polygon area (m ²)	Phase after
X coordinate of the center	Y coordinate of the center	Phase before
Whether the changes in the field.	The class before the changes	The class in database
Width of linear features (m)	The practical classes	The class after the changes
Explanation	The polygons' location, use and description of site.	
The Image block of Polygons.		
Notes: Figure extracted from within the industry's adjustment to the spot and the description of the boundary.		
People fill in a form: Inspected: Filling time:		

Remark 1. Shaded fill in the field, non-shaded need fill within the industry in Table 1.

Polygon number. Organized in the order from left to right, from top to bottom, and the number should be in the field of TBBH without re-numbering. Polygon coding: adopt the word form that head should be in the north and the numerator is polygon number and ownership, denominator should be land number and slope classification, like ab/cd, as a is polygon number, b is ownership, c is polygon coding and d is slope classification.

3.3 Post-Treatment after Verifying the Outside Industry

3.3.1 Verifying the Outside Industry

In order to verify the interpretation of changed polygon, determine the authenticity of polygons of new construction sites, and land types of changed polygon, measure the width of new linear features, supplement the missing polygon, survey the boundary of greatly changing polygon, and then verify the outside industry, and fill the complete monitoring of polygons and industry survey records, shown in Table 2.

Table 2. Field Investigation of Monitoring Land Use Changes

No.	Whether the changes in the field		The class		Across				Remarks
	Year	NO	Before the change	After the change	Monitored	Change	Field	Cause	
1	-		013	122	14.4		14.4		Indoor interpretation of 102, the actual temporary land is the highway construction, and after the completion of highway will be back to the original phase class...
2	-		013	203	0.7		0.7		
3	-		013	122	1.2		1.2		
4	-		043	203	1.2		1.2		
5	-			102	802.9		802.9		
6		-	033	033					
7		-	013	013					
8		-	013	013					
9		-	032	032					
10		-	032	032					
11	-			102	128.7		128.7		
12	-			102	288.5		288.5		
13	-		122	203	1.3		1.3		
14	-		013	203	14.8		14.8		
15	-		032	203	9.1		9.1		
16	-		013	122	6.8		6.8		
17	-		033	203	0.7		0.7		
18	-		013	203	1.3		1.3		
19	-		033	204	64.2		64.2		

3.3.2 Processing Inside

After verifying the changed polygon, complete the vector data and boundary range of the changed polygon, and then convert the format to the standard interchange format of 2008 land use database.

3.4 Updating the Land Use Database

3.4.1 Updating the Land Polygon

Add the changed polygon vector data to the database, and then display with the land polygon. Break up the land polygon layer by changed polygon range, and then combine the appropriate polygon. In this study, the changed polygons are mostly located along the highway, so, it is typically to update the polygon along the highway. In the process of updating the polygon along the highway, the polygon across the highway should be braked up, and the polygon in the same administrative village should be combined together. Code the combining polygon after the maximum number of the polygon number in the same administrative village (TBHH), and then edit the attribution-land coding (DLBM), land name (DLMC), and ownership (QSXZ).

3.4.2 Updating the Linear Features

Add the changed polygon vector data to the database, and then display with the linear feature layer. In this study, the changed polygon land coding are 122, 203 and 204. So, the linear features under the changed polygon should be deleted. And the linear features along the changed polygon range should be trimmed.

4 Conclusion

1. The land use changing information in Xinrong District from 2008 to 2009 has been obtained by remote sensing dynamical monitoring technique. 19 changed polygons

have been interpreted in the satellite images, and after verifying the outside industry, 5 polygons are not changed, the reasons for this can be seen in Tab.2. According to this, remote sensing technique is an effective method to monitor the land use change in large area. While the existing phenomenon of same spectrum for different bodies and different spectrums for same body, the interpretation should be verified in the outside industry.

2. The land use changes for Xinrong District from 2008 to 2009 are caused by the changes of 013 to 102, 033 to 203 and 043 to 204. These have direct relationship with the economic development these years, especially for the new coal line highway which connected to the DaHu highway, which plays an important role in the economic development. And at the same time, all the changes are correspondent to the overall planning of economic development.

References

1. Huang, J.-z.: Research and Application of Land Survey System Based on MVC and Mobile GIS. Wuhan University, Wuhan (2005)
2. Shi, W.-z.: Several advice of land use dynamic monitoring research by remote sensing and GIS technology. *Guangdong Remote Sensing* 6, 48–51 (1990)
3. Mei, A.-x., Peng, W.-l., Qin, Q.-m., Liu, H.-p.: Introduction to Remote Sensing. Higher Education Press, Beijing (2005)
4. Chao, Z.-h., Chen, Q.-g.: Discussion of MODIS imagery geometric precision correction-Taking the Aletai area as an example. *Grassland and Turf*. 2, 35–37 (2005)
5. Xi, W., Xiaoyan, W., Lai, W., et al.: Extraction of land use changing information using middle resolution remote sensing images. *Transactions of the CSAE* 24(supp.2), 85–88 (2008)
6. Land and Natural Resources : Technical Requirements for Land Use Change Survey (Trial) (2003)
7. GB/T21010-2007. Land Use Classification

The Study of GPU-Based Parallel Hilbert Huang Transform

Ningjun Ruan¹, Wen Zhang¹, ShengHui Yu², Kai Xie¹, and HuoQuan Yu¹

¹ School of Electronic Information, Yangtze River University, Jingzhou 434023, China

² BGP Company of CNPC, Zhuozhou 072750, China

Abstract. This paper proposes a 3D volume clipping method based on volume distance field, which is suitable for texture-based volume rendering and exploit per-fragment operation on the graphics hardware to implement clipping. High frame rates are achieved and therefore interactive explorations and clippings of volume data are supported.

Keywords: Seismic data, HHT, Graphical Processing Unit, Parallel Computing.

1 Introduction

As is well known the mass seismic data is nonlinear and non-stationary and that has limited the development of seismic signal processing for a long time. Hilbert-Huang Transform (HHT) [1] is a new signal processing method put forward in 1998 by Huang et al, which has been proved effective to the seismic data processing[2]. In this letter, after a brief description of the HHT, the GHHT algorithm (HHT based on Graphics Processing Unit (GPU)) is present to make the most of the strong computing capability of GPU [3].

2 The Hilbert-Huang Transforms

The Hilbert-Huang transform is based on such hypothesis, that is, every signal can be decomposed into finite intrinsic mode function (IMF), so we can get the instantaneous frequency of the signal. According to the HHT, if we get a signal, the Hilbert transform of the signal can be described as $y_i(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{c_i(\tau)}{t-\tau} d\tau$. Also, we can get the instantaneous amplitude $a_i(t) = \sqrt{c_j^2(t) + y_i^2}$ and instantaneous phase $\theta_i(t) = \arctan[y_i(t)/c_i(t)]$. Then, the instantaneous frequency can be shown as according the definition and the original signal can be expressed as $s(t) = \text{Re}[\sum a_i(t) e^{i\int w_i(t) dt}]$.

3 The Parallel GHHT Algorithm

3.1 Limitations of the Conventional HHT

Though the seismic data can be nonlinear and non-stationary, HHT has been proved effectual to improve seismic reflection data quality effectively [4]. Such as it is, the

classical HHT still has many drawbacks, and some of them can be described as follows:

- (1) If the cubic spline function is not smooth enough, the signal we get from the fitted curve will exceed its steady state value, and that is what we called overshoot phenomenon.
- (2) If enough extreme points cannot be found, the cubic spline curve may cause some errors, so we can no longer get the precise envelop-line.
- (3) When the frequency of two components is similar or the energy of them differ too big (is much different), the HHT cannot separate the two components, and that will lead to the mode mixing phenomenon.

3.2 The Advantages of GHHT

According to the discussion above, the GHHT is presented. In the core of this algorithm, the number of the extreme points is increased in an appropriate way, so we can improve the precision of the curve-fitting and the envelope curve naturally. The key processes of GHHT are described as follows:

- (1) In order to find extreme points, we put forward the sifting standard, which is the product between standard deviation and a coefficient . So the point will be selected if they satisfy the following condition:

$$\frac{1}{N} \left\{ (\sum_{n=1}^N x_{(n-1)}(t))^2 - \sum_{n=1}^N x_n(t)^2 \right\}^{\frac{1}{2}} > \lambda \times \sqrt{\frac{1}{N} \sum_{j=1}^N (x_j - r)^2} \quad (1)$$

- (2) In order to imitate the original signal, we make a fitting curve to pass through the extreme points. If the overall error reaches minimum, the curve can stand for the total trend of the seismic signal.
- (3) We will compare the true value of the signal with the value we get from the fitting curve. If the former is bigger, a minimum value point will be inserted to balance the trend of the overall signal, and vice versa.
- (4) We can get a more precise envelope curve after above steps.

3.3 Parallel Speedup by GPU

In order to improve the processing speed of this algorithm, we use the GT240 GPU to achieve the parallel processing of mass seismic data [5]. Figure 1 describes the thread decomposition for a 3D volume (circled in red) that represents the entire grid being computed. Blocks (marked in green) process a trace of seismic data using parallel threads (marked in orange) that compute a part of the trace. The GHHT will be divided into two parts: the three-dimensional seismic data is divided into small parallel parts and they can be processed on the GPU; on the other hand, the logic control part is still implemented on CPU. The overall pipeline of GHHT is shown in Fig. 2.

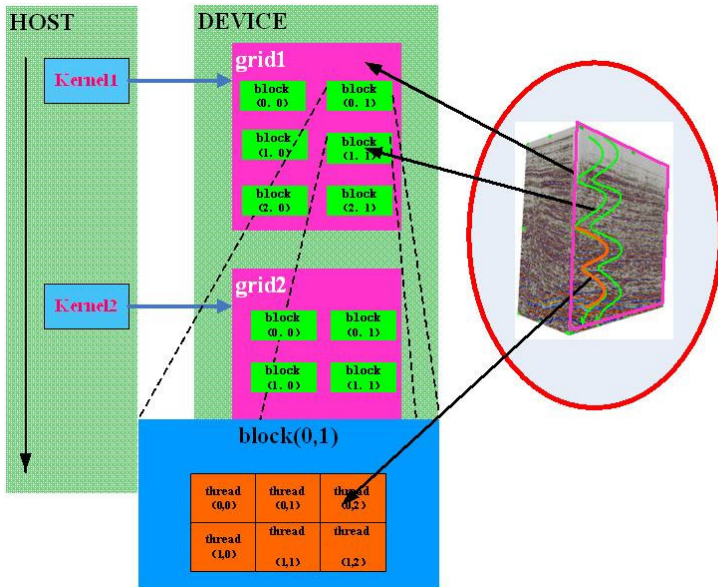


Fig. 1. Thread decomposition on the GPU using NVIDIA's CUDA

4 Experiment Result and Analysis

4.1 The Comparison of Resolution between Original Data and Processed Data

Fig. 3 shows the basic framework of the seismic data processing system developed by our lab, and then the same signal is respectively decomposed by the conventional HHT and the GHHT. On Fig. 4, it is easy to find the GHHT (shown on the right) can decompose the signal into more IMFs (the fourth IMF still contains a lot of useful information while the third IMF obtained from the conventional HHT contains no information nearly), so the GHHT can be used to get the better seismic cross-section. In order to validate our algorithm, we use some single traces to test the resolution of processed seismic data. In the Fig. 5, the GHHT can make more peaks and troughs compare to the original data.

4.2 Parallel Speedup Based on GPU

The GPU has been widely applied to many fields because of its strong computing capability. In our experiments, the GT240 GPU with 12 cores is used to process the seismic data and Fig. 6 shows the speedup ratio with different size of seismic data compare to the CPU processing, we can also find the GHHT obtains the speedup ratio between four and five times.

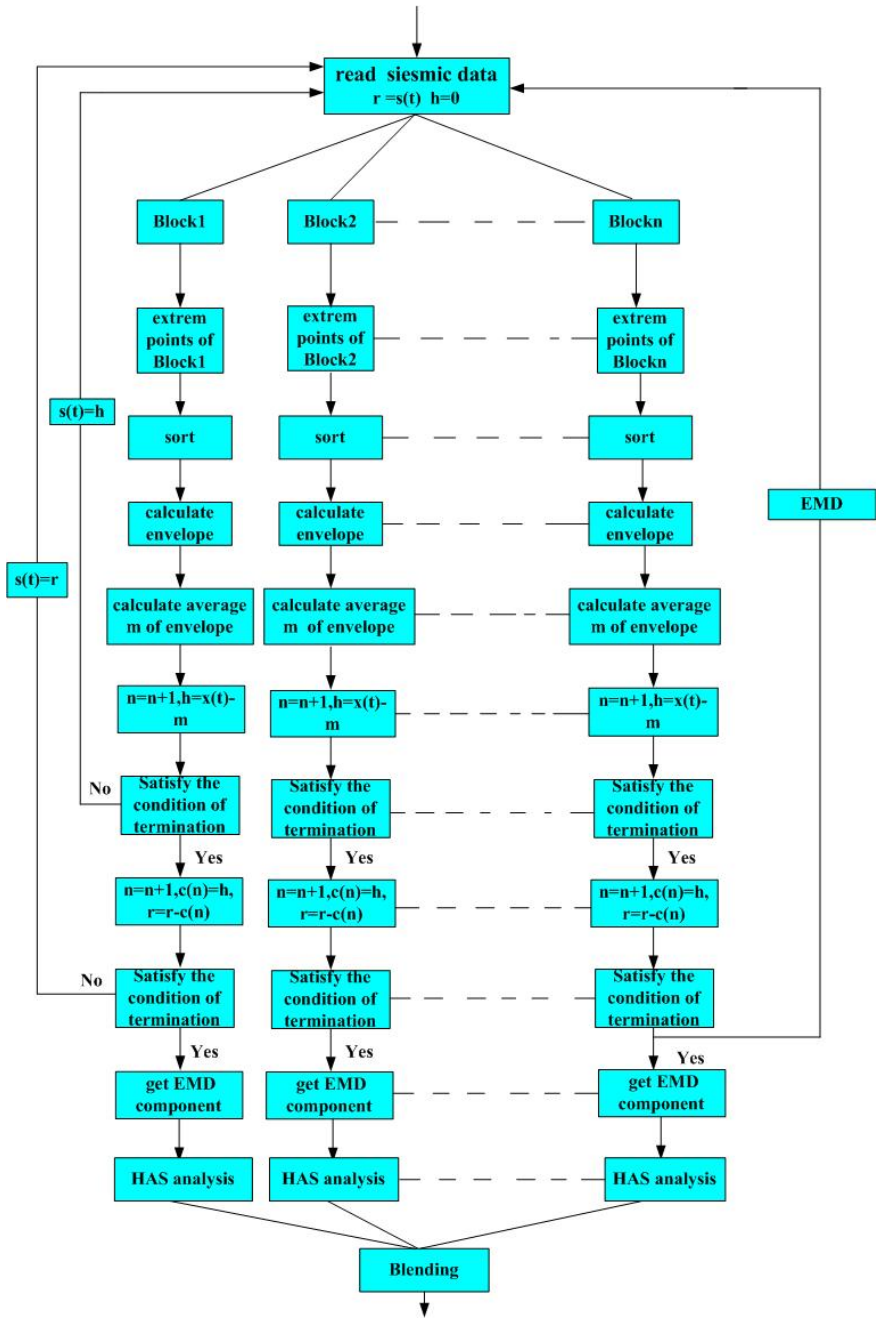


Fig. 2. The overall pipeline of GHHT

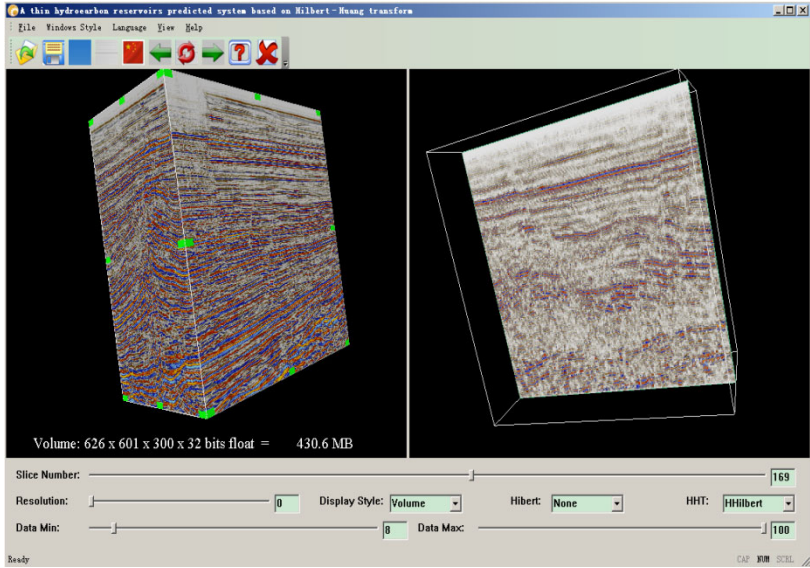


Fig. 3. The basic framework of our system (the original data is shown on the left and the processed data is shown on the right)

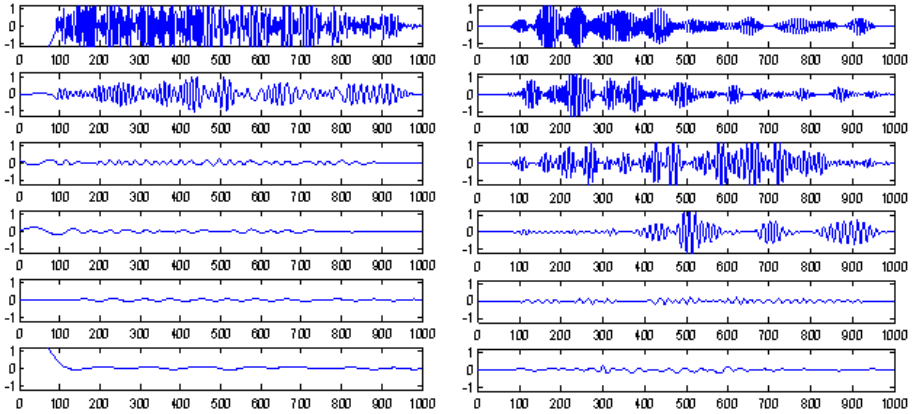


Fig. 4. The comparison of signal decomposition between the original HHT and the GHHT

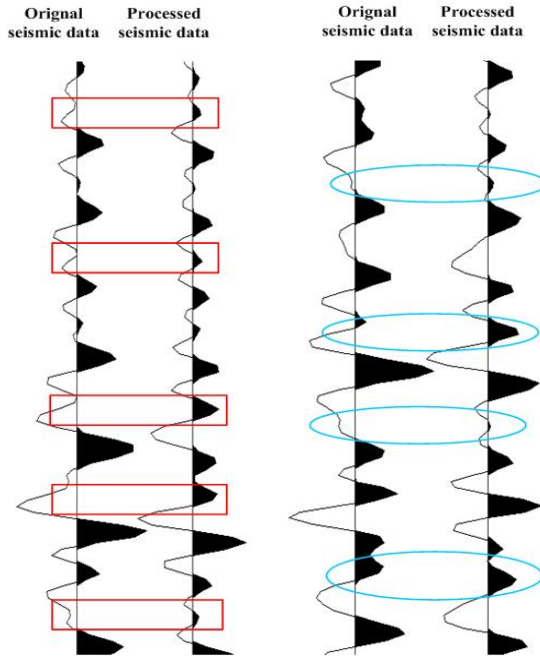


Fig. 5. The comparison of single trace between the original seismic data and the processed seismic data

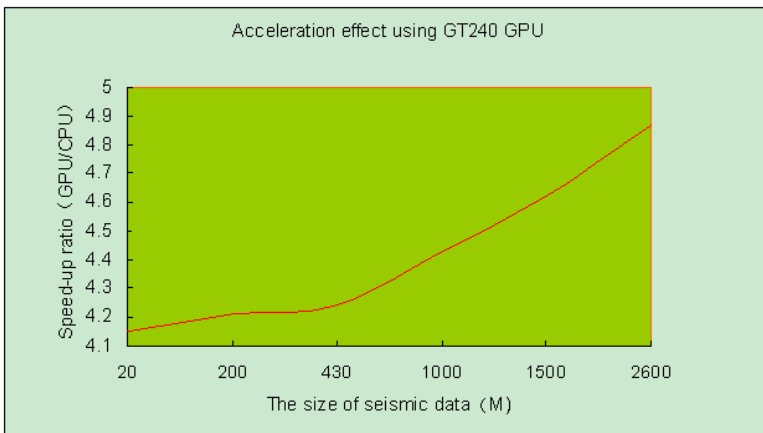


Fig. 6. The speedup ratio by GPU using different size of seismic data

5 Conclusion

A new algorithm called GHHT is put forward to improve the resolution of the stratum and the GPU is utilized to process the mass seismic with its powerful ability for Parallel Processing.

Acknowledgments. This work has been partially supported by CNPC Innovation Foundation (2010D-5006-0304), National Natural Science Foundation of China (60973012, 60973032, 10JJ2045), Natural Science Foundation of Hubei Province of China (2009CDB308), Educational Fund of Hubei Province of China (Q20091211).

References

1. Huang, N.E., Shen, Z., Long, S.R., Wu, M., Shin, H.S., Zheng, Q., et al.: The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R Soc. Lond* 454, 903–995 (1998)
2. Caseiro, P., Fonseca-Pinto, R.: Andrade Screening of obstructive sleep apnea using Hilbert-Huang decomposition of oronasal air pressure recordings. *Medical Engineering & Physics* 24, 1025–1031 (2010)
3. Stone, J.E., Hardy, D.J., Ufimtsev, I.S., Schulten, K.: GPU-accelerated molecular modeling coming of age. *Journal of Molecular Graphics and Modelling* 29, 116–125 (2010)
4. Che, S., Boyer, M., Meng, J., Tarjan, D., Sheaffer, J.W., Skadron, K.: A performance study of general-purpose applications on graphics processors using CUDA. *J. Parallel Distrib. Comput.* 68, 1370–1380 (2008)
5. Huang, N.E., And Shen, S.S.: Hilbert-Huang: Transform And Its Applications. *Interdisciplinary Mathematical Sciences*, vol. 5. World Scientific Publishing Company, Singapore (2005)

Designing and Recognizing Landmark for Positioning Underground Mine Vehicle

Yu Meng^{1,*}, Hongmei Wu¹, Li Liu¹, and Wenhui Li²

¹ School of Mechanical Engineering, University of Science and Technology Beijing, Beijing 100083, China

² College of Computer Science & Technology, Jilin University, Changchun 130012, China
myu_jlu@126.com, flyfirst@263.net, lilu@ustb.edu.cn,
liwenhui2050@163.com

Abstract. Real-time positioning for underground mine vehicle is a key to unmanned underground mining. Machine Vision-based positioning is a novel idea for it. Using Machine Vision method to position devices, landmarks with obvious visual feature and geographic information are needed. In this paper, a type of landmark with barcode is designed and an algorithm for recognizing it is proposed. This type of landmark carries geographic information and can be identified easily. The recognizing algorithm is suitable for real-time mine vehicle positioning because of its small computational and high precision.

Keywords: Unmanned Underground Mining, Machine Vision, Interleaved 2 of 5 Barcode.

1 Introduction

Unmanned underground mining is a way to solve the high risk of the mining industry. For unmanned mining equipments working in narrow tunnels, real-time positioning and navigating accurately technique is needed to avoid the collision between the equipments and wall, direct them moving on correct orbits and ensure unmanned underground mining task working in an orderly way.

Nowadays, there are several types of underground positioning technique, such as Infrared Ray(IR) technique [1], Radio Frequency Identification(RFID) technique [2] [3], Through-the-earth communication technique [4], leaky communication technique [5] and Wireless Local Area Network(WLAN) technique [6]. But neither of them is suitable for unmanned mining independently, because some of them depend on manipulation of human and others can not achieve enough positioning accuracy. Therefore, we propose a Machine Vision-based positioning technique. Its idea is, if the camera on each mine vehicle in tunnel may capture an image with several landmarks placed on given positions at any moment and the geographic information hiding in these landmarks were recognized by image processing method, distances from mine vehicle to these landmarks would also be measured by vision ranging

* Supported by National Natural Science Foundation of China project (No. 50904007).

method so that the positioning result of the mine vehicle could be calculated finally. Machine Vision-based mine vehicle positioning technology does not need manipulation of human. As long as landmark recognition method and ranging method were enough accurate, underground mine vehicle could be located accurately in real-time. Therefore, designing and recognizing landmarks is the first step for Machine Vision-based mine vehicle positioning technology.

2 Designing the Aspect of Landmarks

Landmarks used in underground mine vehicle positioning should have obvious visual feature and be able to represent particular data so that they could carry geographic information and be recognized easily. Therefore, a method that using barcode to make the aspect of landmarks is proposed in this paper.

2.1 Barcode

Barcode technique is developed on the basis of computer technology and information technology, which is the combination of encoding, printing, identification and data acquisition techniques. Graphics of barcode can represent numbers, letters and other symbols. Barcode is composed of bars and spaces with the same width and reflectivity according to certain encoding rules. These crossover distributing bars and spaces have obvious visual feature and can represent data. Therefore, barcode is suitable for making the aspect of landmarks.

2.2 Designing Barcode Landmark

When using Machine Vision-based positioning technique, each image captured by cameras on underground equipments should contain at least one whole barcode landmark so that it can be recognized. Accordingly, the coding rules of barcode landmark should have higher density. We select Interleaved 2 of 5 barcode to make the aspect of landmarks, because all bars and spaces of it represent codes and it has high density.

Interleaved 2 of 5 (also called I 2 of 5) is a numbers-only barcode. The symbol can be as long as necessary to store the encoded data. The code is a high density code that can hold up to 18 digits per inch when printed using a 7.5 mil X dimension. The "Interleaved" part of the name comes from the fact that a digit is encoded in the bars and the next digit is encoded in the spaces. The encoded digits are "Interleaved" together. There are five bars, two of which are wide and five spaces, two of which are wide. The symbol includes a quiet zone, the start character (narrow bar - narrow space - narrow bar - narrow space), the encoded data, the stop character (wide bar - narrow space - narrow bar), and a trailing quiet zone^[7].

The X-dimension is the width of the smallest element in a bar code symbol. The minimum X-dimension for an "open system" (a bar code label that will be read by scanners from outside your company) is 7.5 mils (a mil is 1/1000 inch) or 0.19 mm. The "wide" element is a multiple of the "narrow" element and this multiple must remain the same throughout the symbol. This multiple can range between 2.0 and 3.0 if the narrow element is greater than 20 mils. If the narrow element is less than 20

mils, the ratio must exceed 2.2. Quiet zones must be at least 10X or at least .25 inches.

The height of the bars must be at least .15 times the symbol's length or .25 inches, whichever is larger. The overall length of the symbol is given by the equation:

$$L = (C(2N + 3) + 6 + N)X \tag{1}$$

Where L = length of symbol (not counting quiet zone), C = number of data characters, X = X-dimension, N = wide-to-narrow multiple.

Encoding rules of data are shown in Table 1 and landmarks made according to Interleaved 2 of 5 barcode are shown in fig.1.

Table 1. Encoding rules of Interleaved 2 of 5 data (N denotes Narrow, W denotes Width)

Number	Pattern	Number	Pattern
0	NNWWN	5	WNWNN
1	WNNNW	6	NWWNN
2	NWNNW	7	NNNWW
3	WWNNN	8	WNNWN
4	NNWNW	9	NWNWN

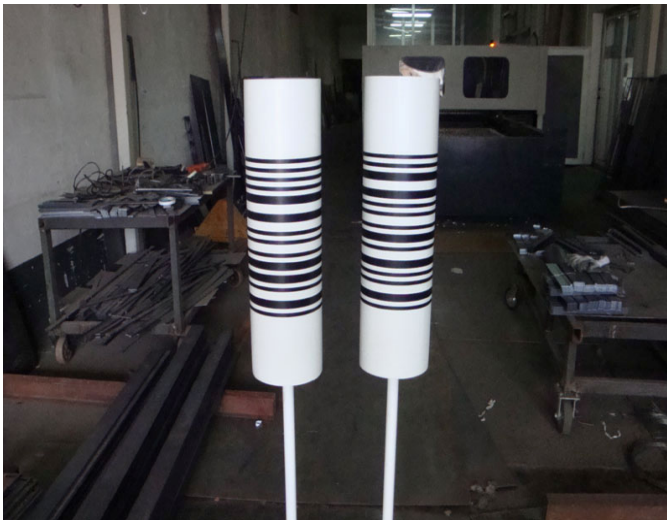


Fig. 1. Landmarks made according to Interleaved 2 of 5 barcode

3 Recognizing Barcode Landmark

Finding the correct scanning direction is the key to recognize the barcodes in images. The most reliable methods for finding the correct scanning direction are to scan the

images along omni-direction, but the computational is too large for real-time applications. Fan^[8] suggested to use moment-invariants theory to rotate barcode images in correct directions, but this method only works for barcode images with simply backgrounds in our experiments. Furthermore, computational of this method is large too.

We propose a barcode recognition method which is similar with the omni-direction scanning method and suitable for recognizing the landmarks made according to Interleaved 2 of 5 barcode. The process of it is as below:

- 1) For the raw image I captured by the camera on an underground device, transform it to image I' by dilation, erosion and smoothing to erase the noises in it. Binarize image I' to image I'' ;
- 2) Denote $L = \{l_i\} (i = 1 \dots n)$ as a pencil of parallel lines in the vertical direction of the image I'' , where l_1 overlaps with the left boundary of I'' , the distance between l_n and the right boundary of I'' is less than 10 pixels and the distance between line l_i and line l_{i+1} is 10 pixels. Set the rotation angle $\theta = 0^\circ$;
- 3) Respectively rotate the pencil L by angle $+\theta$ and $-\theta$, it is cut by image I'' to a clockwise pencil of scanning parallel line segments $L_S = \{l_{S_i}\}$ and an anticlockwise pencil of scanning parallel line segments $L_N = \{l_{N_i}\} (i = 1 \dots n)$. Set variable $j = 1$;
- 4) According to the definition of Interleaved 2 of 5 start character and stop character, judge whether there is start character or stop character on the scan line segment l_{S_j} in image I'' . If there is no start character or stop character, turn step 5. Otherwise, read digits along forward direction or backward direction according to the definition of Interleaved 2 of 5 data. After reading digits, verify the following stop character or start character. If the verification is successful, return the recognizing result and quit the algorithm. If it fails to read digits or verify stop character or start character, scan the next start character or stop character on the line segment until scanning to the end of it;
- 5) According to the definition of Interleaved 2 of 5 start character and stop character, judge whether there is start character or stop character on the scan line segment l_{N_j} in image I'' . If there is no start character or stop character, turn step 6. Otherwise, read digits along forward direction or backward direction according to the definition of Interleaved 2 of 5 data. After reading digits, verify the following stop character or start character. If the verification is successful, return the recognizing result and quit the algorithm. If it fails to read digits or verify stop character or start character, scan the next start character or stop character on the line segment until scanning to the end of it.

6) Set $j = j + 1$, if $j \leq n$, turn step 4;

7) Set $\theta = \theta + 5^\circ$, if $\theta \leq 90^\circ$, turn step 3. Otherwise, judge that there is no barcode in image I'' and quit the algorithm.

Considering the barcode landmarks are placed along the vertical direction to the ground, the correct scanning direction is close to the vertical direction to the ground usually. Therefore, the scanning angle for barcode starts from the vertical direction and rotates gradually along clockwise and anticlockwise direction respectively.

4 Experiments

We place the barcode landmarks on the corridor of our laboratory and take a sample car as a mobile device, on which there are a high-speed camera and a embedded computer. Recognizing result is shown in Table 2 by using the method proposed in this paper.

Table 2. Recognizing result of the method proposed in this paper

Content of Landmark	Recognizing result	Recognizing time (s)
(00,01)	(00,01)	0.087
(35,82)	(35,82)	0.125
(00,01)	(00,01)	0.132
(35,82)	(35,82)	0.114
None	No Landmark	0.332
(35,82)	(35,82)	0.122
None	No Landmark	0.285
(00,01)	(00,01)	0.095
(35,82)	(35,82)	0.106

After repeated experiments, the precision of the landmark recognition algorithm reaches 98.6%. For images with barcodes, the algorithm may recognize the barcode information in 0.1 second approximately; for images without barcodes, the algorithm may return the correct result in 0.3 second approximately. Therefore, this algorithm can satisfy the demand of real-time recognizing landmark.

5 Conclusion

For Machine Vision-based method to position underground mine vehicles, landmarks with obvious visual feature and geographic information are needed. We use barcode to design the aspect of landmarks and propose a barcode recognizing algorithm. This

algorithm has small computational and high recognizing precision. It can judge whether there is barcode in an image or not and recognize barcode (if there is) in a short time. Therefore, this algorithm can satisfy the demand of real-time recognizing landmark and is suitable for underground mine vehicle positioning.

References

1. Luke, R.A.: Diamonds with dispatch: computerized LHD dispatching at DeBeers Finsch- A progress report. *World mining Equipment* 6, 24–28 (1992)
2. Huang, Q., Lu, Y., Sun, Z.: Development of KJ90 Following Location and Attendance Management System for Underground Personnel. *Mining Safety & Environmental Protection* 31(5), 13–16 (2004)
3. Tang, B., Xu, T., Luo, M., He, X.: Radio Frequency Identification Management System for Personnel in the Well Based on nRF403. *Modern Electronics Technique* 29(7), 10–12 (2006)
4. Li, J.: Application of PED System in Coal Mines. *Express Information of Mining Industry* 10, 94–95 (2004)
5. Outalha, S., Le, R., Tardif, P.M.: Toward the unified and digital communication system for underground mines. *CIM Bulletin* 93(1044), 100–105 (2000)
6. Li, J.: Research on the UHF Radio Wave Propagation and Location Algorithms in Mine Tunnels, PHD Thesis, Tianjin University (2006)
7. <http://www.adams1.com/i25code.html>
8. Fan, Y., He, H.: Study of recognition system for 1-D and 2-D bar code image. *Chinese Journal of Scientific Instrument* 24(suppl.4), 488–489 (2003)

Researches on Combustion of Vacuum Switching Arc Based on Image Processing Method

Huajun Dong¹, Yongquan Gan¹, Guiming Shi², and Jiyun Zou³

¹ School of Mechanical Engineering, Dalian Jiaotong University Dalian, China

² School of Institute of Information Engineering Dalian University of Technology Dalian, China

³ Department of Electric and Electronic Engineering Dalian University of Technology Dalian, China

Abstract. Combustion process change of vacuum switching arc and heat radiation induced by combustion process in the vacuum gap have great influence of breaking capacity of vacuum circuit breakers. In this paper a set of image collection system was installed to collect vacuum switching arc images with high-speed and high-resolution CMOS camera. Then digital image processing technology is applied to the research of mutuality of vacuum arc image, according to the diffusion rate of vacuum switching arc in the whole igniting process, whole combustion process of vacuum arc can be divided into five parts: 1) initial stage, 2) diffuse stage I, 3) stable stage, 4) diffuse stage II, 5) quenching stage; the distribution of diffusion flame velocity in different vacuum gaps and different magnetic is gotten at the same time. From the results we can see that in the same conditions of other parameters the greater the vacuum gap, the greater the gradient of internal vapor pressure in arc column and the faster radial diffusion speed of arc is; at the same time, with the controlling of axial magnetic, the breaking capacity of vacuum circuit breakers can be improved greatly. Combined with the results in this paper the understanding of the vacuum arc combustion diffusion flame mechanism is deepened.

Keywords: Vacuum switching arc, CMOS, image processing, MATLAB, diffusion, expansion velocity.

1 Introduction

Vacuum switching metal vapor arc in the regional of arc column consists of neutral atoms, electronic and ion. Because of arc beam internal steam pressure in arc column, charged particles and neutral will exo-proliferate from arc column. This diffusing process has a major impact on the breaking capacity and reliability of vacuum circuit breakers [1, 2]. Therefore many scholars at home and abroad have done many researches on diffusing process of vacuum switching arc using high-speed camera equipments. Heberlein [3] has put forward that the spread appearance of vacuum arc is the necessary conditions of arc fault current through zero, combined with experiments he found several arc shapes: 1) Diffusion state arc beam, 2) Shrink state

arc beam, 3) Eruption arc beam and 4) anode eruption patterns. At the same time, he got the relationship between arc diameter and transient current, arc voltage and arc form. He also analyze the influence of axial magnetic field to arc shape.. Analysis of the speed change of spatial expansion in the vacuum switch arc burning process and the factors which affect its change such as vacuum gap, magnetic field etc can help improve the vacuum circuit breakers' work performance, reliability and their application in high-voltage and high capacity. Changes in image sequences arc burning can be got using high-speed CMOS in this paper, what's more expansion speed of arc flame can be analyze using digital image processing correlation analysis method. With the comparison of metal vapor arc radial expansion speed in different vacuum gaps and in different magnetic fields, the morphological characteristics of expansion in the burning process of vacuum switching arc can be researched.

2 The Related Analysis and Algorithm of Digital Image

Area is a convenient objective measure of the object scope. Area is related to the change of image gray level, it is fully determined by the edge of the area and the objects. In digital image processing techniques, there are several algorithms used to calculate the area: 1. Pixel count method; 2. Boundary travel yards calculation method; 3. Boundary coordinates calculation method. The most simple calculation method is to count the number of pixels in the boundary, find out the sum of the pixels in the domain boundary. Using the pixel count method to count the area of domain, not only simple, but also the best estimate to unbiased and consistent of the original analog area. Its computation formula is [4]:

$$S = \sum_{x=1}^N \sum_{y=1}^M f(x, y) \tag{1}$$

In this formula, $f(x, y)$ means binary image it signify objects when its value is 1, it signify background when its value is 0, Its area is to count the number of $f(x, y) = 1$.

During the actual object area calculating set R_x to signify the horizontal resolution of image (pixel/inch), set R_y to signify the longitudinal resolution (pixel/inch), set s_p to signify the area of single pixel, setting s to signify the area of arc domain setting N to signify the number of pixels in arc area. So the actual area of arc zone is [5]:

$$s_p = \frac{1}{R_x \times R_y} \tag{2}$$

$$s = N \times s_p = \frac{N}{R_x \times R_y} \tag{3}$$

In order to analyze the expansion trend of vacuum arc and simplify the calculation analyze, the number of pixels is used directly to signify the area of arc image. Using

high-speed CMOS camera two consecutive frames arc images can be got in a short time. We can regard the metal vapor spread movement of the burning arc to uniform motion. According to the definition of movement speed may have arc radial expansion speed:

$$V_s = \frac{\Delta S_s}{\Delta t} \tag{4}$$

In this formula ΔS_s : two frames of the arc high displacement vector before and after; Δt : the sampling time interval between two frames.

3 Experimental Results

The mechanism of vacuum switching arc in short-gap is eager to be researched urgently. Fig.1 shows that in the condition of capacitor voltages 80V, the value of vacuum gap is 10mm and 7mm respectively, radial expansion speed of arc changes with time curves. From the figure we can see that in the same conditions the greater the vacuum gap, the greater the gradient of internal vapor pressure in arc column and the faster radial diffusion speed of arc is. When the area of arc reaches maximum the diffusion speed of arc start to become a negative value, at this point, vacuum arc area become smaller and smaller until to zero.

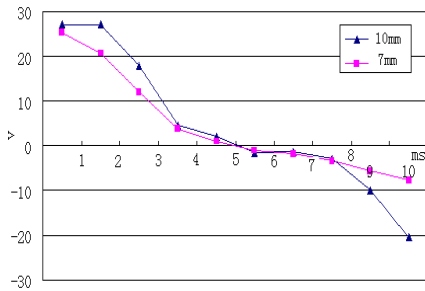


Fig. 1. Radial diffusing speed in different vacuum-gap

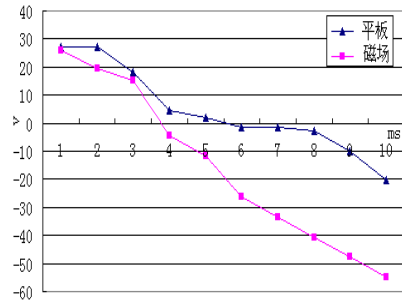


Fig. 2. Radial diffusing speed in different configure contacts

Magnetic controlling technology is often used to control the behavior of vacuum arc to improve the working performance and reliability of vacuum circuit breakers. At present two kinds of magnetic control technology, transverse magnetic field control and axial magnetic control are applied. Application of axial magnetic field can improve the breaking capacity of vacuum switch improved greatly. At present and axial magnetic field has become the main structure of high-current vacuum interrupter chamber contacts. In order to find the control action of axial magnetic to the arc, the voltage between the capacitor is 80, the structure of contacts is flat and 1/3 circle axial magnetic field structure respectively in the condition of vacuum distance is 7mm. Fig2 shows that in arc area increasing stage, because of the effect of magnetic field arc is

restricted in the magnetic field, its radial diffusion speed below the radial diffusion speed of flat contact. The loss of charged particles in arc column is reduced so that it could reduce the evaporation of metal vapor and prevent contact surface from being melted and the breaking performance is improved. However, the arc current reaches the maximum, arc controlled by magnetic reduced rapidly, and it is conducive to the rapid of arc, thus breaking the circuit.

4 Conclusion

The morphological characteristics of vacuum switch arc are studied with digital image processing technology quantitative research of the change law in the combustion process of arc is realized. Experimental results show that: 1) In the combustion process, the height and width of arc current are expanded in certain acceleration. The practical factors of accelerate expansion in the arc burning process maybe: The greater the vacuum gap, resulting in metal vapor within the vapor pressure gradient is greater, thus contributing to the expansion of the vacuum arc movement. Because of the magnetic field control, arc on the electrode surface avoid being constricted, and maintain confusion. 2) This arc behavior of the expansion is symmetrical. The speed of its expansion is determined by the basic combustion speed, arc area, reaction zone and the expansion of hot metal vapor. With the ignition of arc and metal gas expansion, the speed of the wave for the promotion of arc around the move keeps increasing, the same to the size of the structure arc. Thus, with the advance of the arc, its speed increases, and further accelerate the turbulence, because turbulence to increase the reaction area.

References

1. Jimei, W.: Theory and Application of Vacuum switches. Xi'an Jiaotong University Press, Xi'an (1986)
2. Jianwen, W.: Survey on the dynamic image of vacuum arc and electron dissipation process. Xi'an Jiaotong University, Xi'an (1995)
3. Yanqing, W., Huajun, D., Xiaojing, T., et al.: Variations in Area and Shape of Vacuum Switching Arc Images. Chinese Journal of Vacuum Science and Technology 28(6), 531–534 (2008)
4. Dong, H-j., Dong, E-y., Wu, Y-q.: XXIVth Int. Symp. on discharges and electrical insulation in vacuum – Braunschweig, pp. 221–224 (2010)
5. He, L.: Digital Image Processing and Application. China Power Press, Beijing (2005)

A Platform-Based Model for Automatic Planning and Corresponding Complexity Results

Yunpeng Wu, Weiming Zhang, Zhong Liu, Jincai Huang, and Cheng Zhu

College of Information System and Management, National University of Defense Technology
Changsha 410073, China

Abstract. The technology developed from AI planning has been applied in a lot of areas. However, how to model the planning process is a problem that has not been solved yet. In this paper, we model the process from the perspective of platform. A platform is defined as a linear system, by which sequential plan can be obtained. We also propose a parallel planning model based on multi-platforms, and provide a method for parallelizing a given sequential plan. At the last, we define the executable problem of platform-based planning, and analyze the corresponding complexity.

Keywords: Planning, Model, Linear system, Complexity.

1 Introduction

In recent time, planning techniques have been applied in more and more areas, such as robotics, process planning, web-based information gathering, autonomous agents, spacecraft mission control and even military task achievement. However, although planning, as a problem, has been studied for several decades, it is still difficult to be solved completely[16]. A good model can explain the planning process more clearly, thus, can help to study the problem better. State space based planning model formalizes planning conditions with states, and plans with sequence of actions[11][3]. Classical planning model assumes that complete knowledge is obtained before starting the planning process [6][4][9][5]. HTN model decomposes a planning mission into sub-missions which can be achieved directly by single action[8][14]. A lot of works has been done in this area. Baral(etc.)[1] shows that the planning problem in the presence of incompleteness is more difficult than its complete counterpart. Helmert(etc.)[13] provides a map of the computational complexity of non-optimal and optimal planning for the set of domains used in the competitions. Work considers using database technology for automatically planning has been proposed[7]. Blum firstly proposed a graph-based planner which always returns the shortest possible partial-order plan[2]. Bernd(etc.) [15] presents a framework to integrate hierarchical task network planning and hierarchical scheduling approach.

2 Preliminaries

For the sake of simplicity, we firstly introduce several notions that will be used in the sections. A literal l is atomic, denotes a fact in the world. A state q is consisted

represents a configuration of the world. The set of all possible states is called a state situation s is a part of a state, represents a subset of facts which should be focus on. Each literal or state is a situation. An action a is atomic, for each action a , there are two situations corresponds to it. One is the precondition for executing a , the other represents result which is made by executing a . for simply, we call them initial situation and situation, respectively. A platform p is a system which is used to execute actions.

3 Platform as Linear System

Usually, actions those change the initial world to the objective one are the things we mainly focus on. However, platform as the executor of these actions should also be given enough consideration.

Definition 1. A Platform is defined as a tuple $p = \langle L, Q, A, \delta \rangle$, where

- L is a finite set of literals,
- Q is a finite set of states,
- A is a finite set of actions,
- δ is transition function: $Q \times A \rightarrow Q$.

Intuitively, a platform is a linear system, which has the ability to execute different actions. In this paper, we use a tuple $\langle q_I, q_O \rangle$ to represent a planning mission, where q_I is the initial state and q_O is the objective state. Given a platform p and a planning problem instance $\langle q_I, q_O \rangle$, the aim of planning is to select and order a set of actions to change the world from q_I to q_O .

In order to capture the behaviors of planning with a platform, we define a running process of a platform. For a platform $p = \langle L, Q, A, \delta \rangle$, the run of p can be divided into two phases. The first one can be considered as a expanding process of an state tree t . Firstly, t is initialized with the initial state q_I . Consider t constructed so far, for each leaf node q_l , suppose there are totally k actions, in A , can be executed under q_l , then t spawns k successor nodes q_1, \dots, q_k for q_l , each one corresponds to an action. The expansion of t will continue until q_O is achieved or there is no action can be executed under q_l . Obviously, if $b = q_1, \dots, q_l$ be a branch of t which ends at q_l , then q_l can be transformed into q_I through executing corresponding actions by p . The second phase is a synthesizing process, starting from the leaf nodes of t . Let q_l be a leaf node that can not be extended anymore, if $q_l = q_O$, it is labeled with true value, otherwise, it is labeled with false value. For each interval node, it is labeled with true value if there exists one successor node labeled with true value, otherwise, it is labeled with false value.

The result of the running of a platform is a state tree, each node corresponds to an action. If the objective state is not same as the initial state and can be achieved by the platform, then the state tree will contains at least two nodes, one is the root, otherwise, empty. Specially, for each branch which is ended at a state q_l , if q_l is labeled with true, then the actions with total order corresponds to the nodes of the branch is an executable plan.

Example 1. Consider a Block World Problem instance, which has four blocks b_1, b_2, b_3, b_4 in total, and a table T which is used to hold these blocks. The initial and objective states are shown as in Fig. 1. Suppose there is only one platform p can be used. The

running of p is as follows. Firstly, t is initialized with q_1 , actions $a_1 = M(b_3, b_4, b_1)$, $a_2 = M(b_4, b_3, b_2)$, $a_3 = M(b_3, T, b_1)$, $a_4 = M(b_4, T, b_2)$ can be executed under q_1 . t is transformed by spawning 4 successor nodes for q_1 , each one corresponds to an action above. For the first successor node $q_1 = On(b_3, b_4) \wedge On(b_4, b_2) \wedge On(b_2, T) \wedge On(b_1, T) \wedge Clear(b_1) \wedge Clear(b_3)$, and its corresponding action is a_1 . And further, actions $a_{11} = M(b_3, b_1, b_4)$, $a_{12} = M(b_1, b_3, T)$, $a_{13} = M(b_3, T, b_4)$ can be executed under the state q_1 , thus t is transformed by spawning 3 successor nodes for q_1 , each one is corresponding to an action above. Here, the action a_{11} is an action that transforms q_1 to the initial state q_1 . Iteratively process the leaf node of t , the expansion will continues until the terminate condition is satisfied. Consider a branch of t , whose corresponding actions are $a'_1 = M(b_3, T, b_1)$, $a'_2 = M(b_1, b_3, T)$, $a'_3 = M(b_4, T, b_2)$, $a'_4 = M(b_2, b_1, T)$, it is a executable plan, by executing which, q_1 can be transformed into q_0 .

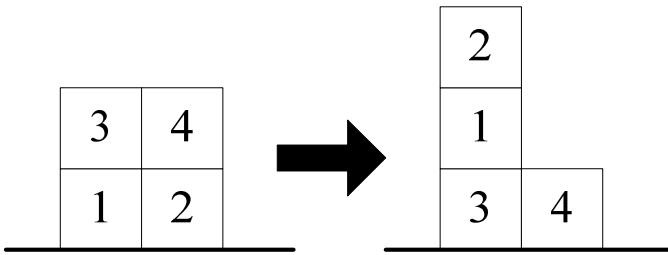


Fig. 1. The initial state and objective state of example 1

Definition 2(Existing problem). For a planning problem over a platform, whether there exists an executable plan.

Theorem 1. The existing problem of planning based on platform is NP-complete.

Proof sketch. We make a reduction from the existing problem over platform to the SATISFIABILITY problem[12]. As the objective state is a propositional expression of literals, and given a state, each literal can be assigned either true value of false value.

Definition 3. Given two states q_1, q_2 and a platform p , the reachability from q_1 to q_2 over p means that q_1 can be transformed into q_2 through executing a sequence of action by p . If p is understood by context, it can be simply called the reachability from q_1 to q_2 .

For a given state q and a platform p , from the result of running p which uses q as the initial state, a map of reachability from q to all other states can be obtained, i.e., the state tree, rooted with q . Especially, given a planning problem instance $\langle q_i, q_0 \rangle$, the reachability from q_i to q_0 can also be obtained from the state tree. If there exists a leaf node labeled with true value, then q_i to q_0 is reachable, i.e., there exists an executable plan.

4 The Parallelization of Sequential Plan

Given a planning problem instance and a platform, all executable plans can be obtained by running the platform, each one of them is consisted as actions with total order.

However, in some times, total order is not necessary. If the sequence plan can be executed in parallel, the total executing time can be saved.

Definition 4. Given a sequential plan, if it has at least two actions that can be executed at the same time, we call it has the parallelizability.

Given a sequential plan, suppose it has k actions, represented as a_1, a_2, \dots, a_k , respectively, and the corresponding initial situations and objective situations are $\langle s_1, s'_1 \rangle, \langle s_2, s'_2 \rangle, \dots, \langle s_k, s'_k \rangle$. Dealing with each $s_i, 1 \leq i \leq k$, for each fact l in s_i , check all objective situations before current one in back-forward order, suppose the current objective situation is s'_j , the plan can be parallelized by using following steps:

- 1) if s'_j satisfy l and $j \neq i$, then check the next objective situation s'_{j-1} ;
- 2) if s'_j does not satisfy l and $j + 1 \neq i$, then make a_i be a successor action of a_{j+1} , and check next fact in s_i ;
- 3) if s'_j satisfies l and $j = i$, also q_i does not satisfy l , then make a_i be a successor action of a_j ;
- 4) if all the facts in s_i have been checked, and a_i does not be a successor of any actions, then make it be an initial action.

After the parallelization by using above steps, unnecessary temporal in the sequence plan will be removed, thus a parallelized plan can be obtained. Intuitively, the parallel plan can be represented by a directed graph, where each node represents an action. Specially, for any two action nodes in the directed graph, if there does not exist a path from one to another, then the two actions can be executed in parallel.

We next propose a parallel planning model to characterize the behaviors of planning with multiple platforms.

Definition 5. A planning problem instance over a set P of platforms (PPM) is defined as a tuple $\langle L, P, Q, A, \theta, \varphi \rangle$, where

- L is a finite set of literals,
- P is a finite set of platforms,
- Q is a finite set of states,
- A is a set of actions,
- θ is a mapping function: $A \rightarrow 2P$,
- φ is a transition function: $Q \times 2A \rightarrow 2Q$.

Here, the mapping function $\theta(a)$ maps each action $a \in A$ to a subset $P' \subseteq P$ of platforms, where each platform $p \in P'$ can execute the action a .

As we may note that given a set P of platforms, it can be seen as a virtual platform p , which can execute all actions that can be executed by platforms of P . All executable sequence plan of p can be obtained by running p , and further, the parallelized plan can be obtained by executing the parallelization steps depicted above. The work is not stop here, since that executing the parallel plan by using a PPM, a mapping from actions of plan to platforms of PPM should be given. Note, given a parallelized plan, it may not have such a mapping, since that the platforms of PPM are fixed, and there will exist conflicts if one platform is mapped to more than two actions that to be executed in parallel.

Definition 6. Given a parallelized plan and a PPM, if there exists a mapping from the actions of the plan to the platforms of the PPM, such that there does not exist conflicts, i.e, one platform is mapped to more than two actions that to be executed in parallel. We say that the parallelized plan is executable by the PPM. If the PPM is understood by context, it can be simply say the parallelized plan is executable.

The executable problem for planning with multiple platforms can be divided into different classes, based on the language used by mapping rule θ . and the complexity of its existence problems are different.

Theorem 2. The executable problem

1. is undecidable when mapping rule is in first order logic.
2. is PSPACE-hard when mapping rule is in the qualified propositional logic.
3. is NP-Complete when mapping rule is in the propositional logic.

Proof sketch

1. Inspired by work[10], we make a reduction from the satisfiability problem for the first order logic querying to the existence problem of PPM. For any satisfiability problem instance $\varphi(I)$, where φ is the query depicted by the first order logic and I is a database instance, there exists a planning instance such that I be the platforms instance and φ be the platform query. Since the satisfiability problem for the first order logic is undecidable, thus the existence problem is undecidable.

2. We make a reduction from AFA to PPM. For each AFA, there exists a PPM such that for each input symbol of AFA there is a literal of L , the disjunction of accepting states of AFA be the initial state q_0 , the initial state of AFA be the target state q_1 , the transition function of AFA be the transition rules.

3. We make a reduction from SAT problem to the existence problem of PPM. Let the objective state q_0 be a propositional formula, and literals in L be the propositional variables, then for each SAT problem instance there is a planning instance, vice versa.

5 Conclusion

In this paper, we define a platform as a linear system, and give the steps to generate sequence plan. Further, we give a method to parallelize a sequence plan, and then we define a parallel planning model based on multiple platforms, and analyze the complexity of its executable problem. Future work should be focus on the follows: first is the optimal problem of planning with platforms. Second, since the executable problem we consider here has high complexity, special cases should be figured out, which would simpler and neater for implications. Third, the model works under the assumption that the observation of the world is obtained completely, but in most applications, the world is partial observed, thus the more analysis should be given on the latter case.

Acknowledgement. This work is supported by the National Natural Science Foundation of China under Grant No.91024006, No.71031007 and No.71001105.

References

1. Baral, C., Kreinovich, V., Trejo, R.: Computational complexity of planning and approximate planning in the presence of incompleteness. *Artificial Intelligence* 122(1-2), 241–267 (2000)
2. Blum, A.L., Furst, M.L.: Fast planning through planning graph analysis. *Artificial Intelligence* 90(1-2), 281–300 (1997)
3. Bonet, B., Geffner, H.: Planning as heuristic search. *Artif. Intell.* 129(1-2), 5–33 (2001)
4. Bylander, T.: Complexity results for planning. In: *IJCAI 1991: Proceedings of the 12th International Joint Conference on Artificial Intelligence*, pp. 274–279 (1991)
5. Bylander, T.: The computational complexity of propositional strips planning. *Artificial Intelligence* 69(1-2), 165–204 (1994)
6. Chapman, D.: Planning for conjunctive goals. *Artificial Intelligence* 32(3), 333–377 (1987)
7. Edelkamp, S.: Planning with pattern databases. In: *Proceedings of The 6th European Conference on Planning (ECP 2001)*, pp. 13–24 (2001)
8. Erol, K., Hendler, J.A., Nau, D.S.: Complexity results for htn planning. *Ann. Math. Artif. Intell.* 18(1), 69–93 (1996)
9. Erol, K., Nau, D.S., Subrahmanian, V.S.: Complexity, decidability and undecidability results for domain-independent planning. *Artif. Intell.* 76(1-2), 75–88 (1995)
10. Fan, W., Geerts, F., Gelade, W., Neven, F., Poggi, A.: Complexity and composition of synthesized web services. In: *PODS 2008: Proceedings of the Twenty-seventh ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, pp. 231–240 (2008)
11. Fikes, R.E., Nilsson, N.J.: Strips: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence* 2(3-4), 189–208 (1971)
12. Garey, M.R., Johnson, D.S.: *Computers and Intractability: A Guide to the Theory of NPCompleteness*. W.H. Freeman, New York (1979)
13. Helmert, M.: Complexity results for standard benchmark domains in planning. *Artificial Intelligence* 143(2), 219–262 (2003)
14. James, K.E., Hendler, J., Nau, D.S.: *Semantics for hierarchical task-network planning* (1994)
15. Schattenberg, B., Biundo, S.: A unifying framework for hybrid planning and scheduling. In: *Freksa, C., Kohlhase, M., Schill, K. (eds.) KI 2006. LNCS (LNAI), vol. 4314*, pp. 361–373. Springer, Heidelberg (2007)
16. Stefik, M.: Planning and meta-planning (molgen: Part 2). *Artificial Intelligence* 16(2), 141–169 (1981)

Research of Emergency Logistics Distribution Routing Optimization Based on Improved Ant Colony Algorithm

Huijie Ding

Suzhou Industrial Park Institute of Services Outsourcing
Suzhou, China
dinghuijie@139.com

Abstract. Emergency relief has characteristics of complexity, urgency, sustainability, technicality, and so on. In this paper a mathematical model to seek the shortest delivery time as the ultimate goal is established based on these characteristics, which is on the core of characteristics with the urgency and consider both the road conditions and on shortage of demand point of relief supplies. The problem of emergency logistics distribution routing optimization is solved by the improved ant colony algorithm—Fish-Swarm Ant Colony Optimization, simulation results show that, compared with basic ant colony algorithm, Fish-Swarm Ant Colony Optimization can find the higher quality to solve the problem of emergency logistics distribution routing optimization.

Keywords. Emergency logistics, routing optimization, Fish-Swarm Ant Colony Optimization.

1 Introduction

In recent years, with frequent occurrence of sudden public safety incidents in our country, emergency logistics was born. Emergency logistics, which purpose is to provide needed emergency supplies, personnel and funds in unexpected natural disasters, sudden public events and military conflicts, is extra-ordinary logistics activities aim to maximize time-efficient and minimize disaster losses [1]. Ali Haghani^[2]described the emergency logistics as network flow problem with multi-item and multi-mode, which is limited by the time window, and gave two solutions. Fiedrich^[3]researched the optimization model of distribution to the number of affected sites and transport resources after the earthquake in the case of time and quantity and quality of resources limited, which aims to minimize the number of deaths. Jae^[4]researched the problem how to schedule vehicles to transport the wounded in the situation of the road network under uncertainty. Fish-Swarm Ant Colony Optimization is used in the problem of emergency logistics distribution routing optimization, in order to achieve the goal of delivery time as short as possible.

I. MODEL OF EMERGENCY LOGISTICS DISTRIBUTION ROUTING OPTIMIZATION

Emergency relief has characteristics of complexity, urgency, sustainability, technicality, and so on. The mathematical model to seek the shortest delivery time as

the ultimate goal is established based on these characteristics, which is on the core of characteristics with the urgency and consider both the road conditions and on shortage of demand point of relief supplies.

To facilitate modeling, make the following assumptions:

- (1) Consider only a single emergency distribution center that location is known, all delivery vehicles start from emergency distribution center.
- (2) Distribution material can be mixed, and material what each relief supplies demand point need does not exceed the maximum load of the vehicle.
- (3) Location and demand of relief supplies demand point are known, requirements of each relief supplies demand point is distributed by only one vehicle.
- (4) Type of the vehicle is single, load of the vehicle is known.
- (5) Vehicle services for each relief supplies demand point, only the case of loading without unloading on the way.
- (6) Average speed of vehicles is known and definite, traveling distance is proportionate to running time of vehicles;
- (7) Due to the impact of disasters, emergency transport vehicles must arrive before time that relief supplies demand points required, otherwise, the road which is impassable is abandoned, because of objective reasons.
- (8) After disaster, some of the roads are destroyed, thereby increasing the traffic of other well-sections and leading to traffic jams. Therefore, changes in traffic flow are described by traffic factor. The greater the traffic factor is, the worse the road conditions, the slower the vehicle speed and the longer the time are.
- (9) Short degree for materials of each relief supplies demand point is different. As the subjective factor, the shorter the materials are, the faster the vehicle speed and the less the time is.
- (10) Emergency vehicles are temporarily deployed to the emergency distribution centers from far and near, after the completion of the distribution task, vehicles do not have to return to emergency distribution center.

$$\min T = \sum_{k=1}^m T_k \tag{1}$$

S.T.

$$T_k = \sum_{i=0}^l \sum_{j=0}^l t_{ij} x_{ijk} \tag{2}$$

$$t_{ij} = \frac{d_{ij}}{v} \times \varphi_{ij} \times \mu_i \tag{3}$$

$$\sum_{i=0}^l g_i y_{ki} \leq q \quad k \in [1, m'] \tag{4}$$

$$\sum_{k=1}^{m'} y_{ki} = 1 \quad i \in [0, l], k \in [1, m'] \tag{5}$$

$$\sum_{i=0}^l \sum_{k=1}^{m'} x_{ijk} = 1 \quad j \in [0, l], k \in [1, m'] \tag{6}$$

$$\sum_{j=0}^l \sum_{k=1}^{m'} x_{ijk} = 1 \quad i \in [0, l], k \in [1, m'] \tag{7}$$

$$\sum_{i=0}^l x_{ijk} = y_{kj} \quad i \in [0, l], k \in [1, m'] \tag{8}$$

$$\sum_{j=0}^l x_{ijk} = y_{ki} \quad i \in [0, l], k \in [1, m'] \tag{9}$$

$$t_{ik} < t_{ei} \tag{10}$$

Where, l is the number of relief supplies demand points; m' is the number of vehicles; d_{ij} is distance between i and j ; T_k is travel time of vehicle k ; t_{ij} is time that vehicles travel from relief supplies demand point i to relief supplies demand point j ; g_i is demand of relief supplies demand point i ; q is maximum load of the vehicle; k is number of the vehicle; v is travel speed of the vehicle; φ_{ij} is traffic factor; μ_i is material tight factor; t_{ik} is time that vehicle k arrives point i ; t_{ei} is latest delivery time that relief supplies demand points are limited.

Emergency distribution center number is 0, relief supplies demand point is 1, 2, 3, ..., l , definition of variable x_{ijk}, y_{ki} is:

$$y_{ki} = \begin{cases} 1, \text{ point } i \text{ is serviced by vehicle } k \\ 0, \text{ other} \end{cases} \tag{11}$$

$$x_{ijk} = \begin{cases} 1, \text{ vehicle } k \text{ travel from } i \text{ to } j \\ 0, \text{ other} \end{cases} \tag{12}$$

Formula (1) is objective function of the model, which aims at the least time; formula (2) is travel time of vehicle k ; formula (3) is needed time that the vehicle travels from relief supplies demand point i to relief supplies demand point j , short degree of each relief supplies demand point and road conditions are considered; formula (4) indicates the total weight of goods loaded does not exceed the maximum load of the vehicle itself for any vehicle in the transport process; formula (5), formula (6) and formula (7) ensure each relief supplies demand point has passed by only one vehicle; formula (8) indicates task of relief supplies demand point j is completed by vehicle k transferring through i ; formula (9) indicates task of relief supplies demand point i is completed by vehicle k transferring through j ; formula (10) is time requirements that materials arrive relief supplies demand points.

II Model is solved by Fish-Swarm Ant Colony Optimization

Step1 let $NC=0$ (NC is iteration), $load_bus=0$ ($load_bus$ is load of the vehicle), $\tau_{ij}(0)=C$ (C is constant), $\tau_{ij}(t)$ is residual pheromone of branch i, j at time t , when $t = 0$, the strength of pheromone on the path is equal, initialize parameters;

Step2 put m ants on to the emergency distribution center, $m = \sum_{i=1}^n b_i(t)$ is the total number of ants, $b_i(t)$ is the number of ants at city i on time t ;

Step3 calculate transition probabilities of ant k ;

$$P_{ij}^k = \begin{cases} \frac{[\tau_{ij}(t)]^\alpha \cdot [\eta_{ij}(t)]^\beta}{\sum_{s \in allowed_k} [\tau_{is}(t)]^\alpha \cdot [\eta_{is}(t)]^\beta} & j \in allowed_k \\ 0 & other \end{cases} \tag{13}$$

Where, $allowed_k = (1,2,..,n) - tabu_k$ is city collection that ant can be selected. $tabu_k$ is taboo table, it records the city which ant k has been passed, used to describe the memory of artificial ants. $\eta_{ij}(t)$ is some heuristic information, $\eta_{ij}(t) = 1/d_{ij}$ α, β reflects the importance of pheromone and heuristic pheromone.

Step 4 according to transition probabilities, select and move to the next city, at the same time, add j to $tabu_k$, check whether the vehicle load;

Step5 calculate crowding factor h_{ij} ;

$$h_{ij}(t) = 1 - \tau_{ij}(t) / \sum_{i \neq j} \tau_{ij} \tag{14}$$

$$\delta(t) = \gamma e^{-bt} \tag{15}$$

Where, $\delta(t)$ is congestion threshold, γ is extremely near level, b is threshold variation coefficient.

Suppose, the ant finds a path through the formula (13), and $h_{ij}(t) > \delta(t)$, it is said that the path to the low level of crowding, then choose this path. Otherwise, it will re-select the other path according to formula (13).

Step6 determine whether it is crowded according to crowding factor. If crowded, continue step 7, otherwise, return to step 3, re-select the next city.

Step7 check whether loads of the vehicle reaches the maximum load. If achieved, return to the emergency distribution center;

Step8 check whether $tabu_k$ is full. If no, return to Step3, otherwise continue to step9;

Step9 calculate the objective function, the current best solution is recorded. Update pheromone according the formula (16);

$$\tau_{ij}(t+1) = (1 - \rho)\tau_{ij}(t) + \Delta\tau_{ij}(t) \tag{16}$$

$$\Delta\tau_{ij}(t) = \sum_{k=1}^m \Delta\tau_{ij}^k(t) \tag{17}$$

Where, ρ is evaporation factor of global pheromone which determine the speed of pheromone evaporation, general value is [0,1] [5], $\Delta\tau_{ij}(t)$ is pheromone increment on

the traveled path ij , initial time, $\Delta\tau_{ij}(t) = 0 \cdot \Delta\tau_{ij}^k(t)$ is section k of ants release pheromones on the edge of ij in the course of travel.

$$\Delta\tau_{ij}^k(t) = \begin{cases} Q/L_k & \text{section } k \text{ of ants pass edge } ij \text{ in the course of this tour} \\ 0 & \text{else} \end{cases} \tag{18}$$

Where, Q is constant, L_k is the length of the loop formed by section k of ants in the course of travel.

Step10 If $NC < NC_{\max}$, $NC = NC + 1$, clear $tabu_k$, return step 2, otherwise, it is in the end.

2 Computer Simulation

Suppose, emergency distribution center coordinate is (142,138), emergency distribution center deploy 3 vehicles to 20 relief supplies demand points for material distribution, loading capacity of each vehicle is 8 tons, the average travel speed of the vehicle is $v=60km/h$. The coordinates of each relief supplies demand point, requirements, end time, unloading time of each relief supplies demand point and short degree of each relief supplies demand point are shown in Table 1. Table 2 is the road coefficient of each relief supplies demand point. Extremely near level $\gamma=0.9$, threshold variation coefficient $b=0.0002$, reference [6] gives the basis for parameter settings of α, β, ρ , obtained after repeated testing, when $\alpha=1, \beta=5, \rho=0.6$, computing results is best. Distance between emergency distribution center and relief supplies demand points or among the various relief supplies demand point uses formula (19).

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{19}$$

Table 1. The coordinates, requirements, end time, unloading time and short degree of each relief supplies demand point

No.	C	R(ton)	ET	UT	SD
1	(120, 90)	0.1	4.5	0.1	0.99
2	(175, 40)	0.4	8.5	0.2	0.98
3	(150, 160)	1.2	2.5	0.5	0.65
4	(190, 36)	1.5	4.0	0.4	0.76
5	(95, 150)	0.8	3.5	0.3	0.97
6	(70, 45)	1.3	4.5	0.5	0.85
7	(100, 165)	1.7	4.5	0.2	0.67
8	(110, 180)	0.6	5.0	0.4	0.83
9	(20, 32)	1.2	8.5	0.1	0.94
10	(74, 15)	0.4	7.0	0.3	0.92
11	(150, 110)	0.9	3.5	0.5	0.86
12	(25, 80)	1.3	6.5	0.4	0.87
13	(110, 18)	1.3	7.5	0.1	0.90
14	(70, 170)	1.9	5.5	0.3	0.73

Table 1. (Continued)

15	(120, 50)	1.7	7.5	0.2	0.69
16	(130, 28)	1.1	6.5	0.2	0.88
17	(140, 55)	1.5	5.5	0.3	0.95
18	(85, 82)	1.6	5.5	0.1	0.78
19	(100, 75)	1.7	5.0	0.4	0.87
20	(40, 100)	1.5	6.0	0.2	0.65

C: coordinates,R: requirements, ET: end time, UT: unloading time, SD: short degree

Fig.1 shows the curve of Ant Colony Optimization, its run routes are:

Vehicle1: emergency distribution center—supplies demand point 3—supplies demand point 7—supplies demand point 8—supplies demand point 14—supplies demand point 15;

Vehicle2: emergency distribution center—supplies demand point 11—supplies demand point 4—supplies demand point 2—supplies demand point 17—supplies demand point 16—supplies demand point 13—supplies demand point 6;

Vehicle3: emergency distribution center—supplies demand point 1—supplies demand point 18—supplies demand point 19—supplies demand point 20—supplies demand point 12—supplies demand point 9—supplies demand point 10;

Fig.2 shows the curve of Fish-Swarm Ant Colony Optimization its run routes are:

Vehicle1: emergency distribution center—supplies demand point 3—supplies demand point 11—supplies demand point 1—supplies demand point 19—supplies demand point 18—supplies demand point 6—supplies demand point 9;

Vehicle2: emergency distribution center—supplies demand point 4—supplies demand point 2—supplies demand point 17—supplies demand point 15—supplies demand point 16—supplies demand point 13—supplies demand point 10;

Vehicle3: emergency distribution center—supplies demand point 5—supplies demand point 7—supplies demand point 8—supplies demand point 14—supplies demand point 20—supplies demand point 12;

Fig.3 is an optimal solution optimization curve with Ant Colony Optimization and Fish-Swarm Ant Colony Optimization under the same operating environment. Simulation results indicate that the ability of ants searching is increased since the introduction of the congestion factor in Fish-Swarm Ant Colony Optimization. Therefore, compared with Ant Colony Optimization, Fish-Swarm Ant Colony Optimization for solving emergency logistics distribution routing optimization problem has more advantages and saves delivery time, so that, life and property of the affected areas can be saved by the greatest extent.

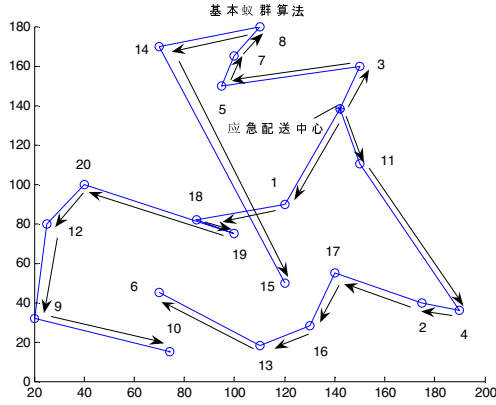


Fig. 1. Curve of Ant Colony Optimization

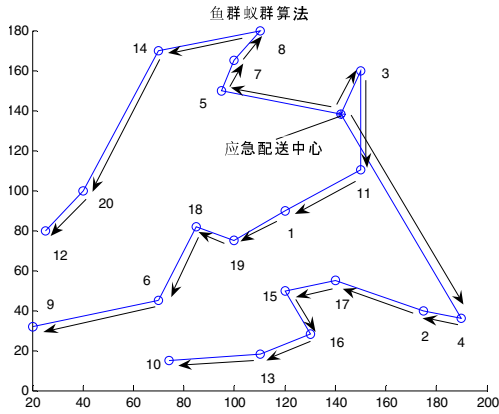


Fig. 2. Curve of Fish-Swarm Ant Colony Optimization

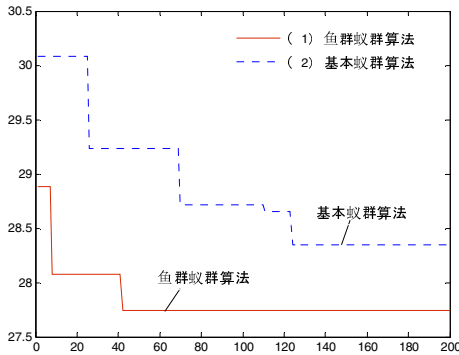


Fig. 3. An optimal solution optimization curve

Table 2. The road coefficient

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	---	1.2	1.1	1.3	1.5	1.6	1.7	1.8	1.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	1.3	1.4
1	1.2	---	1.3	1.4	1.5	1.7	1.9	1.6	1.1	1.1	1.2	1.3	1.0	1.6	1.5	1.4	1.3	1.2	1.0	1.1	1.2
2	1.1	1.3	---	1.5	1.4	1.6	1.7	1.8	1.3	1.2	1.1	1.0	1.8	1.75	1.69	1.74	1.62	1.1	1.2	1.4	1.5
3	1.3	1.4	1.5	---	1.1	1.2	1.3	1.6	1.5	1.3	1.1	1.2	1.4	1.6	1.7	1.81	1.8	1.7	1.9	1.2	1.3
4	1.5	1.5	1.4	1.1	---	1.1	1.2	1.3	1.4	1.5	1.6	1.1	1.7	1.3	1.5	1.6	1.9	1.8	2.1	2.0	1.9
5	1.6	1.7	1.6	1.2	1.1	---	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0	1.9	1.8	1.7	1.6	1.5
6	1.7	1.9	1.7	1.3	1.2	1.1	---	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0	1.9	1.8	1.7	1.6
7	1.8	1.6	1.8	1.6	1.3	1.2	1.1	---	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0	1.9	1.8	1.7
8	1.9	1.1	1.3	1.5	1.4	1.3	1.2	1.1	---	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0	1.9	1.8
9	1.0	1.1	1.2	1.3	1.5	1.4	1.3	1.2	1.1	---	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0	1.9
10	1.1	1.2	1.1	1.1	1.6	1.5	1.4	1.3	1.2	1.1	---	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	1.8
11	1.2	1.3	1.0	1.2	1.1	1.6	1.5	1.4	1.3	1.2	1.1	---	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.7
12	1.3	1.0	1.8	1.4	1.7	1.7	1.6	1.5	1.4	1.3	1.3	1.1	---	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.6
13	1.4	1.6	1.75	1.6	1.3	1.8	1.7	1.6	1.5	1.4	1.3	1.2	1.1	---	1.1	1.2	1.3	1.4	1.5	1.6	1.5
14	1.5	1.5	1.69	1.7	1.5	1.9	1.8	1.7	1.6	1.5	1.4	1.3	1.2	1.1	---	1.1	1.2	1.3	1.4	1.5	1.4
15	1.6	1.4	1.74	1.81	1.6	2.0	1.9	1.8	1.7	1.6	1.5	1.4	1.3	1.2	1.1	---	1.1	1.2	1.3	1.4	1.3
16	1.7	1.3	1.62	1.8	1.9	1.9	2.0	1.9	1.8	1.7	1.6	1.5	1.4	1.3	1.2	1.1	---	1.1	1.2	1.3	1.2
17	1.8	1.2	1.1	1.7	1.8	1.8	1.9	2.0	1.9	1.8	1.7	1.6	1.5	1.4	1.3	1.2	1.1	---	1.1	1.2	1.1
18	1.9	1.0	1.2	1.9	2.1	1.7	1.8	1.9	2.0	1.9	1.8	1.7	1.6	1.5	1.4	1.3	1.2	1.1	---	1.1	1.0
19	1.3	1.1	1.4	1.2	2.0	1.6	1.7	1.8	1.9	2.0	1.9	1.8	1.7	1.6	1.5	1.4	1.3	1.2	1.1	---	0.9
20	1.4	1.2	1.5	1.3	1.9	1.5	1.6	1.7	1.8	1.9	1.8	1.7	1.6	1.5	1.4	1.3	1.2	1.1	1.0	0.9	---

3 Conclusion

In the early stages of solving the model, the congestion threshold of Fish-Swarm Ant Colony Optimization is close to 1, so ants can be free to choose the next city. When the higher pheromone path is confronted with, because of the number of agglomerative are likely more in high pheromone path, it can choose another path. Therefore, disadvantage trapped in local optimal of Ant Colony Optimization can be avoided. Fish-Swarm Ant Colony Optimization can enhance the global search ability, seek the higher quality solutions to solve the problem of emergency logistics distribution routing optimization, and open up a new solution for emergency logistics optimal route selection.

References

- [1] Ou, Z., Wang, H., Jiang, D., et al.: Emergency Logistics. Journal of Chongqing University(Natural Science Edition) 27(3), 164–167 (2004)
- [2] Ali, H., Serchang, O.: Formulation and Solution of a Multi- Commodity Multi- Modal Network Flow Model for Disaster Relief Operations. Transportation Research Part A 30(2), 231–250 (1996)
- [3] Fiedrich, F., Gehbauer, F., Rickers, U.: Optimized Resource Allocation for Emergency Response After Earthquak. Disasters Safety Science 35(1), 41–57 (2000)
- [4] Jae:Dorctor.Stochastic Scheduling Problems for Minimizing Tardy Jobs with Application to Emergency Vehicle Dispatching on Unreliable Road Networks. University of New York (2003)
- [5] Lidong, L.: Master.Research on Improved Ant Colony Optimization.Southwest Jiaotong University (2005)
- [6] Ye, Z., Zheng, Z.: Study on the parameters in Ant colony algorithm——An example to TSP. Wuhan University (Information Science) 29(7), 597–601 (2004)

A WSN-Based Pulse Wave Velocity Detection System for Arteriosclerosis Monitoring

Deng Chen and Shao Chen

Dept. of Electrical and Information Engineering, Shanghai University of Engineering,
LongTeng Road 333, Shanghai, China
dc@sues.edu.cn, shao1204@163.com

Abstract. Arterial pulse wave velocity (PWV) measurement is of great theoretical importance and practical value in evaluating the arterial elasticity and stiffness. The goal of our work is to focus on health-related applications of wireless sensor network (WSN). In this paper we detail a hardware and software solution of pulse waves detection system based on wireless sense network. It can acquire remote multichannel pulse wave signal. Because of its features such as small, portable, low-power, and wireless transmission, it makes a remote health care system possible between hospital, community health centres and families. The experimental result shows that the system can capture pulse wave signal accurately in real time. The physiologic signal included in signals contribute greatly to detect arteriosclerosis, with practical application value.

Keywords: Wireless Sensor Network, pulse wave velocity, remote wireless monitoring, arteriosclerosis.

1 Introduction

Pulse wave velocity has become a well-accepted surrogate measure of arterial stiffness and has been shown to be a significant predictor of cardiovascular risk in diseased and older healthy populations[1]. Through the pulse waves detection system, we can synchronously detect and record a group of related pulse wave signal by several pressure sensors on the arterial pulse, so as to conduct the meaningful research about some disease such as cardiovascular, hypertension, arteriosclerosis and so on. Along with the social aging's aggravating, long-term chronic illness's guardianship becomes the important social question. Therefore application the wireless network for medical monitoring has the widespread demand[2][3].

Since the 21st century, the wireless sensor network characterized by self-organization, non-center, low power has served as the future communication technology development priorities, and obtained more and more uses in national economy and the people's livelihood. Wireless sensor network adopts Zigbee protocol, which is one kind of wireless technology of short distance (0-75 meters) and low rate (<250Kbps). The amount of data of human body pulse wave signal is not big. If calculating according to the 1KHz sampling rate, the data are about 1Kbps[4]. On

the other hand, ZigBee’s electric current can meet handheld devices need for low power,because electric current in an active mode is less than 20mA, but in an sleep mode is 3-40μA . Aim at the application for remote healthy guardianship, we carry out some researches to designs a pulse waves detection system based on the wireless sensor network. The system acquires pulse wave signal through wirless sensor detection node ,and real time sends the pulse wave signal to monitoring center. Through these signals, monitoring center.can analyse the monitor’s health status[5].

2 System Architecture

The hardware structure of this system is shown in figure 1, mainly consisting of several wirless sensor nodes and relay nodes, gateway nodes and PC (monitoring centre).

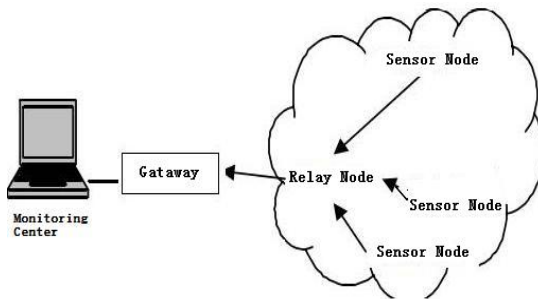


Fig. 1. System Architecture

Wireless sensor nodes mainly finish human body’s pulse wave signal acquisition , signal conditioning , A/D conversion and wireless signal transmission. Relay nodes analysis received packets, and implement data fusion, and wireless send data . to gataway nodes.

Gataway node . transmits received signals to PC via USB conection line. A system software is developed using Labview to process and analyze the sampled data , thus getting a detecting result.

In order to reduce power loss, the technology of hardware low power design and software energy-saving are melt into the process of designing wireless sensoe nodes[6][7]. Figure 2 is the architecture of the WSNs node which is designed in this paper. Figure 2(a) is a scheme of wireless sensor detecting node which is composed of detecting module,processor module, wireless data transeiver and power manage module. Figure2(b) is a block of relay node or gataway node which is composed of processor module, wireless data transeiver, power manage module and communication module.

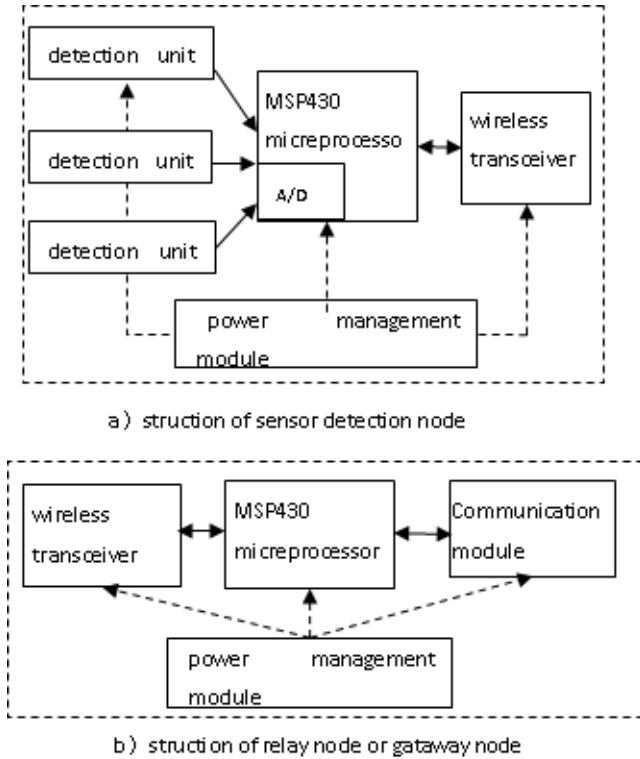


Fig. 2. Block diagram of node

2.1 Detection Unit

The detecting unit which will sample the human body’s pulse wave signals and transform them into electrical signals suitable for A/D conversion mainly contains pulse wave sensor and signal conditioning circuit. In this design, PVDF piezoelectric sensor is used as pulse wave sensor. Compared with other sensors PVDF, piezoelectric sensor has some advantages such as big piezoelectric coefficient, wide frequency response, good match for acoustic reactance, high mechanical strength, good flexibility, lighter in weight, impact resistance, easy to form membrane, lower in price and so on. On the other hand, it can accurately detect human body’s weak pulse wave signals and transform them into electrical signals.

The pulse wave signal from high-impedance’s human body is a kind of weak signals whose signal-noise ratio is low[8]. There is some interference such as baseline drift, human breath, muscle tension and so on, it is 50Hz power interference as well when signal is acquired. In order to get pulse wave signal suitable for processing, the signal must be amplified and filtered. So instrumentation amplifier AD623 produced by AD company is selected as system preamplifier. It has low noise, small zero drift, high input impedance and CMRR. In addition, the

second-order low filter and notch filter are designed to eliminate the interference of human body myoelectricity and breath, as well as 50Hz power interference.

2.2 Microprocessor

Microcontroller (MCU) is the core of wireless sensor network node. It controls and coordinates the work of detection units and wireless transceiver, finishes single processing and signal processing and storage. A lot of MCU produced by TI, ATMEL, Motorola, Renesas and 32 bit ARM are used in WSN nodes. Especially, MSP430 series MCU produced by TI are featured by its ultra low power, low voltage, strong signal processing capability and a lot of peripheral interfaces supporting. In this study, MSP430F1611 is chosen.

Microprocessor suitable for wireless sensor network nodes should have some features such as small size, higher integrated level, low power, support a sleep mode, fast operation speed, abundant I/O interfaces and communication interfaces. Now microprocessor produced by TI, ATMEL, Motorola, Renesas and 32 bit ARM are widely used in wireless sensor network nodes. Especially, MSP430 series MCU produced by TI are featured by ultra low power, low voltage, strong signal processing capability and a lot of peripheral interfaces supporting. In this study, MSP430F1611 is chosen. It has a 12 bit, 250kbps A/D converter, SPI and UART on chip. So it is easy to decrease cost and product sizes.

2.3 Wireless Transmission Module

Wireless transmission module used in this scheme is based on Chipcon CC2420. CC2420 is a transceiver designed specifically for industrial applications in the 2.4 GHz band. It is the RF-IC transceiver compliant with the IEEE 802.15.4 standard and the effective data rate is up to 250 kbps. The multi-point to multi-point network can be fast implemented. It can ensure the effectiveness and reliability of short range communication with little components. It is featured by its delicateness, cheapness and low power. It is suitable to be powered by battery. Moreover, it is also featured by hardware encryption, reliability, flexible networking and high destroy-resistance[9].

It is very easy to connect CC2420 with microprocessor. Figure 3 is the interface connection diagram. Four pins SFD, FIFO, FIFOP and CCA denote the state of receiving and sending data. Microprocessor communicates with CC2420 via SPI interface.

3 Software Design

The function of software system is divided into two parts: program of wireless sensor network nodes and program of monitoring server. The former runs in each wireless sensor network node, the latter runs in PC. The development of software system adopts several development tool platforms. IAR Embedded, Workbench are used for writing program, debugging and writing flash of wireless sensor network nodes. Labview is used for developing monitoring software. So the development of software system is a synthetic and complex project.

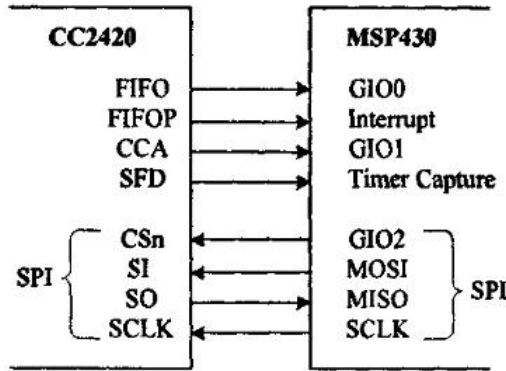


Fig. 3. Interface between MCU and wireless transceiver

The program of wireless sensor network nodes mainly consist of data acquisition module , command and response module and data transmission module.

In data acquisition module , there are programs for data sampling , analog to digital converting , data transmitting. In order to achieve PWV calculations, different positions of pulse wave signals need to be acquired simultaneously by several pressure sensors. In this design ,the radial artery, femoral artery, dorsum pedis artery are selected as the sites for pulse wave acquiring . They are acquired,by use of 8 channels,12 bits A/D integrated into MSP430 and sampling frequency is 1kHz . Wireless transceiver module finishes to pack and transmit the pulse wave signals according to IEEE802.15.4 frame format[9].

In command and response module , PC sends control commands to wireless sensor network nodes ,selects a wireless sensor network node required to communication , commands it to send the pulse wave signals or prohibit other sensor nodes to send information. By default each wireless sensor network node is in non-work model to save power.

Data transmission module is mainly used for wireless data transmission between sensor network nodes and relay nodes , and wire data transmission between a gateway node and PC. Transceiver used in this system is CC2420 , so data frame format transmitted between sensor nodes should meet IEEE802.15.4 standard frame format.. The data transmission between MCU and CC2420 is mainly carried out via a SPI interface. In order to improve the efficiency of data transmission, the simplified format of data segment is shown as figure 4. Wireless sensor network nodes only send pulse wave acquired to a gateway node ,thus the data-packet format only includes ID number of node and the signal to be acquired,as shown in Figure 4(a). The gateway node aggregates the data to be sent from different sensor nodes and passes to the PC. The data-packet format of gateway node is shown in Figure 4(b).

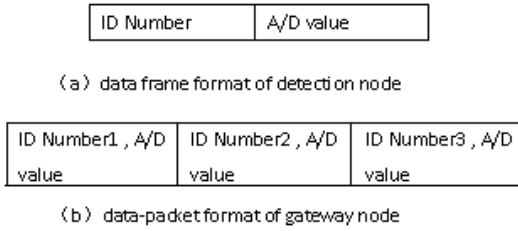


Fig. 4. Data-packet format of data segment

The scheme used for wireless communication is interrupting-mode for wireless receiving and scanning-mode for wireless transmitting.

Wire data transmission is achieved via a UART interface. Table1. is the data-packet format of wire data transmission defined in this system. In which “<” is as packet header and “>” is as packet trailer. ADC_DATA records the pulse wave datas collected from channel A0,A1 and A2. TIME records current time .The system uses 6MHz high frequency crystal as the clock source for baud rate generator. To ensure that gateway node can high-speed transmit data to PC , the register UBR is set to 0x0006., UxMCTL is set to 0x55 and baud rate is set to 115200.

Table 1. Data Packet Format of Wire Communication

header	data	Time	trailer
“<”	ADC_DATA	TIME	“>”

Using PC acts as a monitoring center in this sysytem. The design of monitoring program selects Labview by NI. So monitoring program mainly achieves the following functions: data read and write from serial interface, data analysis, pulse waveform display, waveform preservation of historical data and PWV calculation.

4 Experimental Results

In the present study, the PWV of radial artery, femoral artery and foot arteries were measured in real time. The three waves acquired from detection system of pulse wave main interface in PC are shown in figure 5. New method based on chaotic oscillator for measuring pulse wave velocity [10] in the research project was applied to detect the PWV of carotid artery - radial artery in 10 subjects, and 100 times reproducibility tests were also carried out.Table2. is the test results of the means and variances of PWV. The test results are shown:

- The system achieved non-invasive continuous aquisition of pulse wave signal in normal operating conditions and had some features such as small size, low power and work long hours.
- The variance was far less . Obviously the results of multiple measurement for PWV with this system had no significant difference, and it showed better reproducibility.

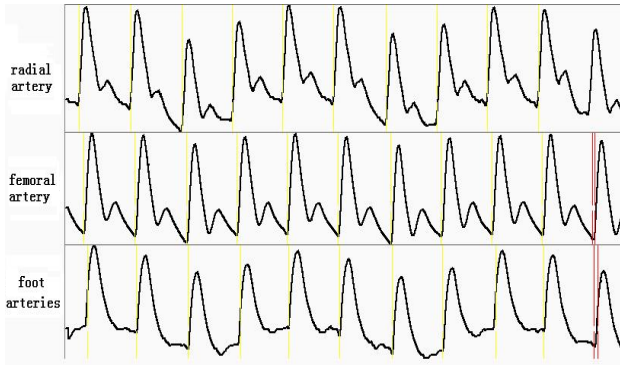


Fig. 5. Detection System of Pulse Wave Main Interface

Table 2. Test Results of CRPWV

subjects	means	variances
1	8.24	0.012
2	9.10	0.015
3	8.03	0.10
4	8.13	0.021
5	7.01	0.018
6	7.30	0.014
7	8.79	0.023
8	9.05	0.018
9	7.83	0.015
10	8.35	0.035

- Normal PWV values of a person is general in the 7 ~ 9m / s, so the pulse wave measurement system based on WSNs designed in the paper can represent PWV value correctly.

5 Conclusion

1. In the present study we design and implement a new type of pulse wave signal wireless detection system based on the principle of wireless sensor networks. Experimental results show that the system can easily detect different parts of the body pulse wave signal at the same time . The system uses wireless communication technology, eliminates the need for complex wiring and has good market prospects.

2. Multiple sensor network nodes can flexibly constitute a wireless network. The network performance is stability. Monitoring and management software can effective monitor network and process the pulse wave signals to be collected .It provides a feasible solution for wireless sensor network-based remote medical monitoring system.

References

1. Qasem, A., Avolio, A.: Determination of aortic pulse wave velocity from waveform decomposition of the central aortic pressure pulse. *J. Hypertension* 51, 188–195 (2008)
2. Baker, C. R., et al.: *Wireless Sensor Networks for Home Health Care*. In: 21st International Conference on Advanced Information Networking and Applications Workshops, pp. 832–837 (2007)
3. Victor, S., et al.: *Sensor Networks for Medical Care.*, Harvard University Technical Report TR-08-05 (2005)
4. Yuanen, J., Wu, X.: The design and implementation of arteriosclerosis examination system based on PWV measurement. *J. Automation & Instrumentation* 133(5), 6–8 (2007) (in Chinese)
5. Peng, Z., Ming, C.: A remote health care system based on wireless sensor networks. In: 12th International Conference on Computer Supported Cooperative Work in Design, pp. 1102–1106 (2008)
6. Zhigang, L., et al.: Transmission Power Control for Wireless Sensor Networks. In: International Conference on Wireless Communications, Networking and Mobile Computing (WiCom 2007), pp. 2596–2599 (2007)
7. Chen, D., YiPing, C., et al.: A WSN-Based Experiment Platform for Monitoring. In: 2th International Conference on Information Technology and Computer Science, pp. 431–434 (2010)
8. Jing, P., Cheng-lin, P.: Application of medical signal processing based on theory and method of chaos. *International Journal of Biomedical Engineering* 29(2), 124–127 (2006) (in Chinese)
9. Chipcon, A. S.: CC2420 Preliminary Datasheet (rev 1.2). Copyright Each Manufacturing Comoany (2004)
10. Chen, D., Wan-qing, S.: New method for pulse wave velocity measurement based on chaotic oscillator. *Computer Engineering and Applications* 46(32), 232–235 (2010)

Condition Assessment Based on Gray Clustering with Cloud Whiten Function for Power Transformer

Zheng Ruirui, Zhao Jiyin, Li Min, and Wu Baochun*

College of Information & Communication Engineering, Dalian Nationalities University,
Dalian, China
zhengrui@yaho.com.cn, limin@dlnu.edu.cn,
wubaochun2007@yaho.com.cn

Abstract. A new gray clustering analysis algorithm, which adopted cloud model as whiten function, was proposed to solve the subjective problem in power transformer condition assessment. The algorithm determined condition classification and parameters through their prior knowledge. The algorithm first optimized indicators of power transformer condition evaluation and stratified them. Then, cloud model was introduced as whiten function of gray clustering because the combination of these two methods could comprehensively consider randomness, ambiguity and gray of uncertainty problems. The parameters of cloud models determined according to prior knowledge of power transformer condition evaluation indicators, so it reflected the power transformer condition more really and objectively. Improved weight coefficient gray target theory analyzed dissolved gases and improved the accuracy of the model. Indicators polarities of the improved weight coefficient gray target theory were optimized so that it could evaluate normal state. For samples only with dissolved gases data, improved weight coefficient gray target algorithm can assess condition alone. Example indicates that condition evaluation results of the power transformer based on gray clustering with cloud whiten function consistent with the actual situation, and raise the scientific and objective of power transformer condition evaluation.

Keywords: Power transformer, condition evaluation, gray clustering, cloud model, improved weight coefficient gray target theory.

1 Introduction

Power transformer is one of the key apparatuses in the electric power system, and its running condition affects the safety and reliability of the electric power grid. Thus, it is significant to correctly access power transformer operating condition [1]. Condition based maintenance for power transformer examines and integrates various pieces of information about power transformer state, and then realizes appropriate condition evaluation for power transformer. Because the maintenance items, frequency and contents of power transformer are determined according to former condition evaluation

* Corresponding author.

results, the maintenance is highly targeted and timely. Thus, power transformer condition maintenance becomes an inevitable trend [2]. Condition assessment for power transformer is the basis and the primary step of its condition maintenance. Running conditions of power transformers are constrained randomly by various factors, some of the factors are uncertainty, so it is very complex to model. Ref [1] introduced weighted gray target theory to insulated condition assessment for power transformer, and gained satisfactory effect. But the condition indices of Ref [1] are not complete. Ref [3] adopted hierarchical gray evaluation model to power transformer condition assessment, and selected appropriate indices. The results of an example confirm the feasibility of the proposed algorithm in Ref [3]. Ref [4] considered both fuzziness and randomness of indices of power transformer condition, and then put forward a condition assessment algorithm based on the cloud model, and its performance proved the cloud model was effective for power transformer condition assessment. Former 3 methods have same disadvantage, which is subjective effect of their condition indices graded strategy. In order to reduce the subjective effect, we introduce the priori statistic knowledge about every index of power transformer condition and use the cloud model as the whiten function of the gray clustering analysis, which combines statistic and random properties of the indices, to improve accuracy of condition assessment for power transformer.

2 Indices and Grade Method of Condition Assessment for Power Transformer

2.1 Indices

It is important to determine the appropriate indices of condition assessment for power transformer. Indices selected must truly and fully reflect the real condition of power transformer, and be easy to obtain. The author of the paper carefully select 4 main indices, they are gases dissolved in transformer oil, electric test, transformer oil characteristics and qualitative indices. Among them, electric test, transformer oil characteristics and qualitative indices have their own sub-indices as shown in Fig. 1. Considering qualitative indices, rather than too many details of each index of former researches, we take environment, history, maintenance and accessories as a whole, and give the experts' scores on them, which simplify the assessments of qualitative indices. In Section 4, the specific assessment method for each index will be discussed.

2.2 Grade Method

There is no standard power transformer condition assessment grade method. Researchers have various grade strategies [5]. In earlier studies, power transformer condition comes in 2 varieties, normal and abnormal. Grade methods become more and more complex and reliable along with the development of study becoming more deeply. But those grade methods always use the same intervals for each grade; meanwhile, the probability of each grade is equal, which is obviously not realistic and objective for assessment. Ref [6] studies 299 power transformers with faults, and presents their statistic distribution of faults serious degrees, which is shown in Table 1. According to Table 1, this paper proposes a 5-level grade method shown in Table 2.

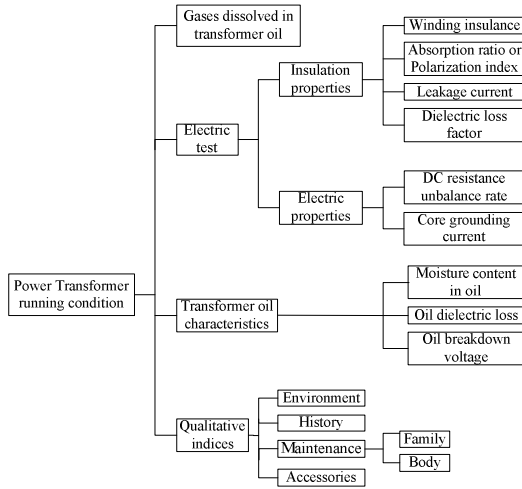


Fig. 1. Power transformer condition assessment indices

Table 1. Distribution of transformer fault serious degrees

Fault serious degrees	The most serious	More serious	serious	Less serious	Total
Number	11	96	140	52	299
Percentage (%)	3.7	32.1	46.8	17.4	100

Table 2. 5-level grade method

Level	Description
1	The most serious fault, stop the power transformer immediately!
2	More serious fault, please arrange its maintenance as possible as quickly.
3	Serious fault, please monitor the power transformer, and arrange its maintenance.
4	Less serious fault, please keep monitoring.
5	The power transformer is normal, may extend the maintenance period.

3 Gray Clustering Analysis with Cloud Model Whiten Function

3.1 Gray Clustering

Gray clustering analysis [7] classifies observed indices or objectives into certain gray categories based on their whiten values of the whiten functions. Suppose the problem has N clustering objects, denoted by $i=1, 2, \dots, N$; S gray categories, denoted by $k=1, 2, \dots, S$; M indices, denoted by $j=1, 2, \dots, M$. Let d_{ij} represent the sample of object i to index j . Then, the sample matrix D is like (1).

$$D=(d_{ij})_{N \times M} \tag{1}$$

Let F be a mapping, $OPf_{jk}(d_{ij})$ is the operation of k gray categories of the j th index for sample d_{ij} , and f_{jk} is the whiten function of the j th index for the k th gray category, e.g. $F: OPf_{jk}(d_{ij}) \rightarrow \sigma_{jk} \in [0, 1]$, where $1 \leq i \leq N$, $1 \leq j \leq M$, $1 \leq k \leq S$, and F is called as gray clustering.

Define the weight of the j th index to the k th gray category is η_{kj} , and then the clustering coefficient of the i th clustering object to the k th gray category is like (2).

$$\sigma_{ik} = \sum_{j=1}^M f_{jk}(d_{ij}) \cdot \eta_{kj} \tag{2}$$

Where, $\sigma_i = (\sigma_{i1}, \sigma_{i2}, \dots, \sigma_{iS})$ is the clustering vector of σ_{ik} , and the corresponding gray category of the maximum σ among σ_i is the category of the clustering object .

3.2 Cloud Model

Gray clustering analysis adopts whiten function to determine the clustering result, so the chosen of the whiten function is quite important for it. Commonly used white function for gray clustering analysis is the triangle function and trapezoidal function. Both of the functions need researcher to determine their boundaries, thus, losing uncertainty and objectivity of gray clustering analysis. Using cloud model as the whiten function of gray clustering analysis can automatically determine its boundary, and combine randomness, fuzziness and grayness of uncertain data. Suppose the problem to be solved has N clustering objects, S gray categories and M clustering indices. Firstly, cloud model C_{jk} of index j to class k is determined. Secondly, X factor cloud calculates the expectation value y_0 of index j to category k . Thirdly, calculating each expectation value of all index and category for clustering object i . Finally, the category corresponding to the maximum is the result. A cloud model has 3 parameters, E_x , E_n and H_c . E_x is the expectation, which is the centroid of all the cloud drops, and is the optimum number that reflects the qualitative concept. E_n is the entropy that represents the accepted data range. H_c is a constant that could be adjusted according to specific index. For an index with bilateral constraints $[C_{min}, C_{max}]$, E_x and E_n are calculated through (3).

$$E_x = (C_{min} + C_{max}) / 2, E_n = (C_{max} - C_{min}) / 6 \tag{3}$$

4 Cloud Models of Indices

DL/T596-1996 “Preventive test code for electric power equipment” [8] prescribes note values for most indices, but there are no clear classifications for different conditions. This paper determines parameters of cloud models not only based on standard DL/T596-1996 but also prior knowledge from other references, thus makes whiten functions in this paper are more objective. As shown in Fig. 1, there are 15 indices; we take only a few typical examples of the indices due to the space limitation.

4.1 Polarization Indices

Polarization index denoted as PI. Standard DL/T596-1996 indicates that PI is normally larger than 1.5. Ref [9] provides Table 3 which reflects the relationship of PI and power transformer condition. Let “danger”, “no good”, “questionable”, “less good” and “good” in Table3 correspond to level 1-5 in Table 2 respectively, and then the parameter for each condition is determined as Table 4 shown in Fig 2.

Table 3. Relationship of PI & power transformer condition

Power transformer condition	PI
danger	≤ 1.0
no good	$1.0 < \sim \leq 1.1$
questionable	$1.1 < \sim \leq 1.25$
less good	$1.25 < \sim < 2.0$
good	≥ 2.0

Table 4. Parameters of cloud model for PI

Condition levels	Parameters of cloud models			Figure
	E_x	E_n	H_e	
level-1	1.0000	0.0333	0.0050	Fig3 (a)
level-2	1.0500	0.0167	0.0050	Fig3 (b)
level-3	1.1750	0.0250	0.0050	Fig3 (c)
level-4	1.6250	0.1250	0.0050	Fig3 (d)
level-5	2.0000	0.1250	0.0050	Fig3 (e)

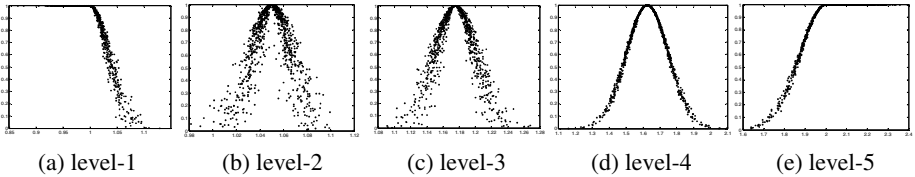


Fig. 2. Cloud models for PI

4.2 Winding Insulance

Standard DL/T596-1996 doesn't indicate specific insulance value. After analyzing Ref [9] and Ref [10], we conclude that at 20°C, the winding insulance $\geq 800 \text{ M}\Omega$ means the power transformer is normal; if the winding insulance $\geq 1000 \text{ M}\Omega$ means the power transformer is in good condition. DL/T572-95 “Operation specification for power transformer” [11] points out: if the winding insulance reduces by 30% compared to the last test at the same temperature, then it is abnormal. Integrated prior knowledge, cloud models for winding insulance of this paper are shown as Table 5.

Table 5. Parameters of cloud models for winding insulation

Condition levels	Parameters of cloud models		
	E_x	E_n	H_e
level-1	0.0300	0.0428	0.0050
level-2	0.1590	0.0428	0.0050
level-3	0.4740	0.0623	0.0050
level-4	0.7305	0.0232	0.0050
level-5	1.0000	0.0667	0.0050

4.3 Gases Dissolved in Oil

Artificial neural network, Gray clustering and support vector machine [3] are used to judge the relationship of gases dissolved in transformer oil and power transformer condition; however, these are subjective methods. Gray target theory [1] can automatically determine power transformer condition degree through gases dissolved in oil, thus it is more objective than other methods. This paper improves the index polarity of Ref [1] by setting its index to be medium rather than maximum. In other words, the normal condition is the standard center of gray target. The corresponding standard pattern of gray target theory is $\omega_0 = (46.1, 21.5, 61.5, 15.8, 1.2)$. The gray target classification is shown as Table 6.

Table 6. Gray target classification based on gases dissolved in oil

Condition levels	Range of gray target degree
level-1	(0.3333, 0.3450]
level-2	(0.3450, 0.4467]
level-3	(0.4467, 0.5949]
level-4	(0.5949, 0.6500]
level-5	(0.6500, 1.0000]

4.4 Qualitative Indices

Operating environment, history, maintenance and accessories are determined by experts on power transformer condition assessment [5]; the equation of membership equals to n_e divide Ne . Where, n_e is the number of experts who consider index i belongs to condition level j ; Ne is the total number of experts.

4.5 Indices Weights

Former analysis indicates that different indices of power transformer condition have different effects on condition assessment. So this paper proposes weight of each index, shown in Table 7. Where, the number in bracket is the weight of each index.

Table 7. Power transformer condition assessment indicators' weights

	First layer	Second layer	Third layer
Power transformer condition assessment indices	Electric test (0.2703)	Insulation properties (0.5000)	Winding insulation(0.1550)
			Absorption ratio or Polarization index(0.1826)
	Gases dissolved in transformer oil (0.3350)	Electric properties (0.5000)	Leakage current(0.2762)
			Dielectric loss factor(0.3861)
			DC resistance unbalance rate(0.6217)
			Core grounding current(0.3783)
	Transformer oil characteristics (0.2857)	H ₂ (0.2197) CH ₄ (0.1652) C ₂ H ₆ (0.2622) C ₂ H ₄ (0.1525) C ₂ H ₂ (0.2004)	Moisture content in oil(0.4206)
			Oil dielectric loss(0.3703)
			Oil breakdown voltage(0.2091)
	Qualitative indices (0.1090)	Environment(0.1038) Running History(0.2871) Maintenance(0.1807) Accessories (0.4285)	Family(0.4723)
Body(0.5277)			

5 Example Analysis and Conclusions

Example: the test recorders of a power transformer SFSZ9-31500/110 are as follows [5]: H₂=166μL/L, CH₄=28μL/L, C₂H₆=9μL/L, C₂H₄=11.3μL/L, C₂H₂=0μL/L, Moisture content in oil=21mg/L, Oil dielectric loss=2.75%, Oil breakdown voltage=50.1kV, Leakage current=32μA, Absorption ratio=1.41, tanδ=0.341%, DC resistance unbalance rate = 1.19%, Core grounding current=37.4mA.

Analysis is from the third layer to the first layer. Gray target of gases dissolved in oil is 0.4633, which has the probability of 0.0666 for level 3. The white value of transformer oil characteristics has the probability of 0.2105 for level 5; has the probability of 0.6781 for level 4. Because the white value of insulation properties is 0.5430 belonging to level 5 and the white value of electric properties is 1 belonging to level 5, the electric test has the probability of 0.7715 belonging to level 5. So, for the first layer, power transformer has the probability of 0.3016 belonging to level 5; has the probability of 0.0223 belonging to level 3.

Integrated assessment conclusion: this power transformer is in normal condition that is level 5; maybe don't have maintenance this period. The real condition of the power transform is normal, which matches our assessment and superior to Ref [4] and Ref [5].

The condition assessment algorithm and 5 level method for power transformer proposed in this paper fully consider the prior knowledge of each condition index, thus reduce the subjective effect of other methods, and promote higher accuracy and objectiveness. The algorithm also has practice value for other assessment problems in other fields.

Acknowledgement. This work was financially supported by the Fundamental Research Funds for the Central Universities (DC10010103), Fundamental Research Funds of Department of Education of Liaoning Province (L2010094), Talented Faculty Funds of Dalian Nationalities University (No.20116203).

References

1. Ruirui, Z., Jiyin, Z., Baochun, W., Jianpo, L.: Method for Insulative Condition Classification Evaluation of Power Transformer Based on Weight Coefficient Grey Target Theory. *Transactions of China Electrotechnical Society* 23, 60–66 (2008)
2. Hang, J.: Study on Approach to Synthetically Assessing Transformer Condition. North China Electric Power University, Baoding (2005)
3. Hao, X., Caixin, S., Jun, Z.: A Hierarchical Grey Evaluation Model for Operation Condition of Power Transformers. *Automation of Electric Power Systems* 31, 55–60 (2007)
4. Lei, H., Liping, L.: Synthetic evaluation of power transformer condition based on cloud model. *Journal of North China Electric Power University* 36, 98–103 (2009)
5. Ruijin, L., Qian, W., Sijia, L.: Condition Assessment Model for Power Transformer in Service Based on Fuzzy Synthetic Evaluation. *Automation of Electric Power Systems* 32, 70–75 (2008)
6. 胡勇,程蕾. 大型电力变压器故障实例统计分析. *Electric Safety Technology* 5, 20–22 (2003)
7. Zheng, R.-r., Zhao, J.-y., Wang, Z.-n., Wu, B.-c.: Power transformer fault diagnosis based on improved gray clustering analysis. *Journal of Jilin University (Engineering and Technology Edition)* 38, 1237–1241 (2008)
8. DL/T596-1996 Preventive test code for electric power equipment, power industry ministry of Republic of China (1996)
9. Qian, W.: Study of the Comprehensive Assessment Method for the Power Transformer Condition in Service with Fuzzy Theory. Chongqing University, Chongqing (2005)
10. Zhihua, C., Jianguang, Z.: Application of Condition Based Maintenance Technology and its Computer Aided Analysis System. *Power System Technology* 27, 16–24 (2003)
11. DL/T572-95 Operation specification for power transformer, power industry ministry of Republic of China (1996)

Study on Intelligence Detection of Methane Gas Based on NDIR Technology

Yong-zhi Liang¹, Xing-fu Rong¹, Bing Li¹ and Chun-tao Liu²

¹ Institute of Mechanical Eng., Taiyuan University of Technology,
79 west yingze street, Taiyuan, China

² Dept. of Chemical & Environment Eng., Beijing Institute of Technology,
5 South Zhongguancun Street, Haidian District, Beijing, China
{Liangyongzhi, Libing}@tyut.edu.cn, rqs5619@263.com
Liuchuntao@bit.edu.cn

Abstract. In this paper, with regards to the disadvantages of traditional gas sensor used in mine, an intelligent infrared detection method for methane gas based on Non-dispersive Infrared technology is presented. Through the introduction of the principle of infrared spectrum absorption, further to the adoptions of differential measurement method and the technology of single light beam & dual wavelength detection, the relationships between pyroelectric detector signal and gas concentration has been analyzed in details. A relevant formula has been given for between the gas absorption ratio F_a and the peak value of pyroelectric detector terminal voltage. Experimental results show that the measurement error is less than $\pm 0.1\%$ (when concentration of methane gas $< 4\%$), and meet the requirement of coal mine safety operation.

Keywords: Intelligence detection, *NDIR* gas sensor, Infrared absorption.

1 Introduction

In a safe operation of the coal mine, it is a key method to prevent the gas explosion in a coal mine by monitoring and controlling the gas concentration in a reliable, accurate and real time manner. For the time being at a domestic coal mine, it is to utilize the combustion catalyst-type gas sensor in term of gaseous detection, the disadvantage is the shortage of its service life, and frequent calibration is required with higher cost for maintenance, and poisoning occurs easily. At the same time, there are a numbers of problems existing subject to the effect of environment, such as poor reliability and lower accuracy[1] [2]. Therefore, it is of substantially importance to study and develop a novel gas sensor.

This article is to detect the gas concentration utilizing the absorptive feature of methane gas towards infrared spectrum in accordance with infrared spectrum absorptive principle, there are a little impact by the site conditions, and the differential measuring method is adopted to restrain the zero point drift. In comparison to the conventional combustion catalyst type gas sensor, it has significant features for its higher reliability and accuracy. It also features for a number of merits for a long calibration cycle with strong anti-interference capability. It is a novel intelligent gas detection system.

2 Theoretical Description of Infrared Spectroscopy

The principle of infrared spectrum absorption indicates a varied absorption peaks due to variations of vibration and rotation under the effect of spectrum for a specific compound. A specific gas has a varied absorption spectrum against infrared radiation. A featured spectrum absorption intensity of a specific gas has a relation with the concentration of said gas. By way of this principle it can be used to detect the concentration of a specific gas to measure the absorption spectrum to identify the gas of this kind, and to measure the absorption intensity to identify the concentration of gas. Each substance has its own specific absorption spectrum, methane(CH_4) molecule has a four inherent vibration modes, and accordingly a four base frequencies would be generated, the length of wave are respectively 3.433, 6.522, 3.312 and 7.658 μm [3]. Therefore, the concentration of gas can be detected according to the variation of absorption peaks at some specific wave length along the spectrum curves for varied gases.

Refer to fig. 1 below indicating the infrared absorption spectrum of methane gas. It can be seen that absorption intensity of methane gas at central infrared zone exceeds the same at near infrared zone significantly. [4] Therefore, the wave length at 3.39 μm of methane gas is chosen in this article to detect the gas concentration.

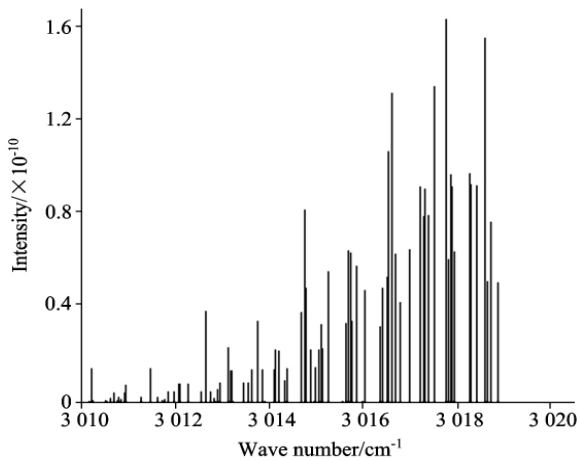


Fig. 1. Absorption spectrum of methane gas near 3.39 μm

When an infrared light passes through the detected gas, its remitted light intensity is in compliance with Lambert-Beer absorption law[4]. In order to eliminate interference generated as a result of the fluctuation of light source in the light path, in which, a dual wave length differential measurement technology is adopted. That is to introduce the dual window detectors. One is an active detector (targeted gas is absorbed), another one is a reference detector (targeted gas is not absorbed). The absorption wave length and reference wave length of the detected gas would be achieved by adding an interference light filter at the upstream of the detector window. As for the methane gas, active detector window would use the interference light filter whose central wave

length is $3.39\mu m$. Reference detector would use the interference light filter whose central wave length is $4.0\mu m$. Ratio between active detector signal and reference detector can be taken to determine the concentration of gas, and the impact on the detector signal that may be resulted from factors including light source and environmental changes would be eliminated[5].

The relation between signals from two detectors and targeted gas concentration is:

$$I = I_0 \exp(-\tau l C) \tag{1}$$

Where: I is a signal in proportional to the active detector, I_0 is a signal in proportional to reference detector, τ is a absorption factor of targeted gas, C is concentration of a targeted gas, l is length of light path to detector from the light source, i.e. the length of absorption gas chamber.

Absorption factor of targeted gas is expressed as:

$$1 - I/(Z \cdot I_0) = 1 - Act/(Z \cdot Ref) \tag{2}$$

Where: Z (Zero) is the I/I_0 value in the reference gas (such as nitrogen), which being a zero calibration value. Act and Ref indicates respectively the signals from active detector and reference detector.

The relation between absorption factor and gas concentration is:

$$1 - I/(Z \cdot I_0) = s \cdot (1 - \exp(-\alpha \cdot C)) \tag{3}$$

given from formula (2) as

$$1 - Act/(Z \cdot Ref) = s \cdot (1 - \exp(-\alpha \cdot C)) \tag{4}$$

Where: $\alpha = \tau \cdot l$ is a constant, s (span) is I/I_0 value of a targeted gas at full concentration, which being calibrated value at full scale stored in *EEPROM* that would be changed when a second calibration is needed. It is certain that temperature is a factor for consideration in the process of calibration that a compensation measures be appropriately taken[6].

Where: the detected gas concentration would be expressed as below given through formula (3) and (4):

$$C = (-\ln(1 - (1 - Act/(Z \cdot Ref))/s)/\alpha) \tag{5}$$

In a practical application, the value of targeted gas concentration is achievable through the study on the ratio of output voltage signal peak to peak value of the detector. Hence, the absorption percentage F_a is defined as:

$$F_a = 1 - S_1/R \cdot S_2 \tag{6}$$

Where; S_1 and S_2 refer to a peak to peak value of the output voltages in respective of detector 1 (measurement terminal) and 2 (reference terminal); R is defined as:

$$R = S'_1/S'_2 \tag{7}$$

Where: S'_1 and S'_2 refer to a peak to peak value of the output voltages in respective of measurement detector and reference detector when the concentration of methane

gas is zero. In a practical measurement, the pure nitrogen is commonly used to measure the S_2^r and S_2^t .

3 Experimental Detection System Setup

In a experimental detection system in this article, a thermal- release electrical detector *TPS2534* is adopted, and a pulsed infrared radiation with *IRL715* be served as light source, the length of its wave range would have a coverage up to $4.4 \mu\text{m}$ from visible light . Refer to Fig. 2 below illustrating a driven circuit of pulsed light source. The circuit is to drive the *MOSFET* by mean of a square wave emitted from the micro-processor, with a 100Ω current limiting resistance connected in series way in a circuit. The driven voltage for the light is 5 V (operating scope is $3\text{-}5 \text{ V}$), frequency is 4 Hz , and duty ratio is 50% .

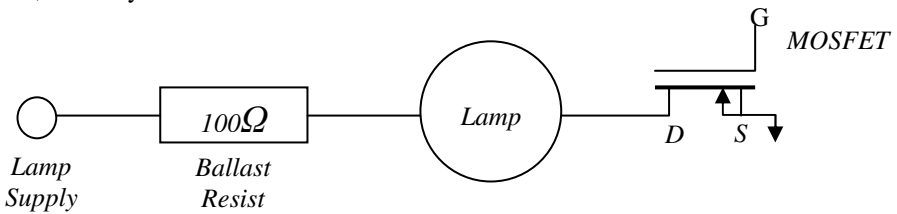


Fig. 2. Lamp power supply arrangement

The detected light beam that enters into the absorption gas chamber would be absorbed by a detected gas, while the reference beam would not be absorbed. Two light beams would be converting to a voltage output in relation to concentration through a detector and a signal-conditioning circuit[7] [8]. The experimental system principle block diagram of methane gas is shown in Fig. 3.

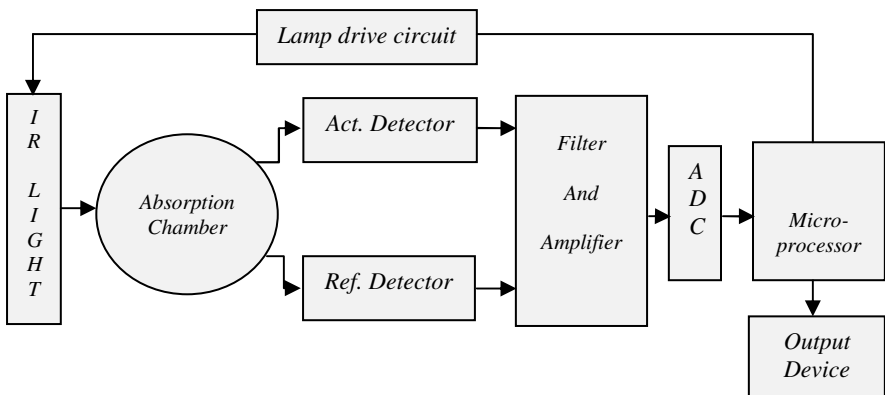


Fig. 3. System principle block diagram

The system would adopt the operating mode with a differential measurement technology and single light beam and a dual wavelength incorporated. Under the effect of a 4 Hz pulsed light source, the measurement detector and reference detector would generate an output signal that is made proportional to the concentration of methane gas under the coal mine. After having been processed through a pre-amplifying wave filter circuit, it would be converted to a digital signal through A/D converter, single-chip microprocessor would indicate an readout for the A/D converted outcome. The value of gas concentration would be achieved following a specific computation, and to realize the functions including the demodulation, display of gas concentration and alarming via a master computer. A multiple infrared detectors would form a detection network covering the whole mine when they are connected through cables to achieve detection for multiple points.

4 Results and Discussion

The experiment is conducted in a closed vessel, a GXH-1050 methane infrared gas analyzer is adopted to calibrate the gas concentration. At a 20 ± 3 °C ambient temperature condition, a high-purity nitrogen would be firstly used to flush the closed vessel, followed by injecting sample methane gas whose purity is 99.9% , and to take notes of the output concentration in full range (0~100%) after the methane gas is injected. The experimental data is shown in Table1.

Table 1. Experimental data of a detector at ambient temperature 20 ± 3 °C

No.	Real concentration/%	Measured concentration/%	Output voltage/V
1	0.5	0.51	3.85
2	1.0	1.01	3.82
3	2.0	2.01	3.78
4	3.0	2.99	3.73
5	3.5	3.49	3.70
6	10.0	10.05	3.64
7	30.0	30.01	3.25
8	50.0	50.21	2.81
9	70.0	70.19	2.46
10	90.0	90.26	2.13

As what is indicated above that the output voltage would decrease along with the increase of the gas concentration basically in a linearity relation. The result of the experiment proves that the error between the concentration value measured by the detector and the actual value would be less than $\pm 0.1\%$ (when concentration of methane gas $< 4\%$), which is of a higher measuring accuracy to meet the requirement for detection of gas concentration in a coal mine. Fig.4 indicates the comparison between the measurement concentration and the real gas concentration.

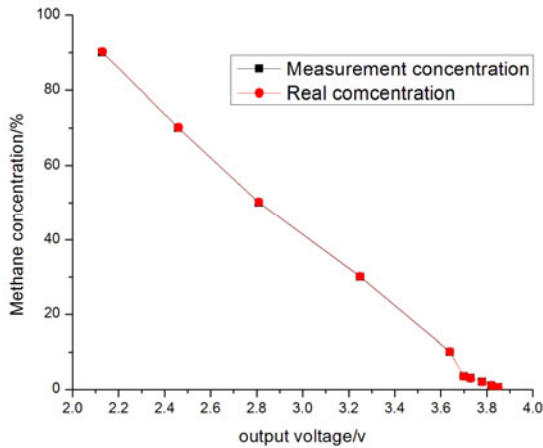


Fig. 4. The comparison between the measurement concentration and the real gas concentration

5 Conclusion

On the theoretical basis of infrared spectrum absorption, the intelligent system of gas detection based on *NDIR* technology was designed for methane detection in coal mine. In this detection system, an operating mode in single light beam with dual wavelength is adopted, and the technique of square wave pulsed modulation for infrared light source is used, in which, the interference of noise is restrained effectively, and the system structure is simplified. Furthermore, a differential measurement method is adopted to eliminate the systematic errors and environmental disturbance. Experimental outcome indicated that the detection accuracy (error $< \pm 0.1\%$ when CH_4 concentration $< 4\%$) in a more stable and reliable operating conditions is suitable for applications in a coal mine.

Acknowledgments. The authors gratefully acknowledge the Research Project Supported by Shanxi Scholarship Council of China. The Project No.2011-030.

References

1. Cheng, Y., Zhang, Y.F.: The design of a new type of high-reliability methane sensor. *Industrial Instrumentation & Automation* 45(4), 69–71 (2006)
2. Yuwei, Y., Zhao, S., Wang, R.: Research of new type of intelligent infrared absorption gas sensor used in mine. *Industry and Mine Automation* (2), 28–31 (2008)
3. Luo, D.-f., Yang, J.-h., Zhong, C.-g., et al.: *Spectroscopy and Spectral Analysis* 31(2), 384–386 (2011)
4. Zhang, L., Yin, W.-b., Dong, L., et al.: *Journal of Optoelectronics Laser* 5(18), 154 (2007)

5. Mao, X., Chen, T., Luo, Y.: Study on new non-poisoning infrared methane detection system. *Chinese Journal of Scientific Instrument* 29(4) (April 2008)
6. Yi, S.H., Park, J.S., Park, J.M.: Temperature compensation of novel NDIR CO₂ gas sensor. *Dig. of IEEE Sensors*, 1373–1376 (2006)
7. Lu Zh, Q., Wang, R.L.: Gas detection with infrared absorption principle. *Coal Science and Technology* (1), 21–23 (2005)
8. Pan, Z., Ma, S., Sun, B., Jiang, X., et al.: *Electronic Measurement Technology* 31(12), 113–115 (2008)

Research on Extracting Semantic Orientation of Chinese Text Based on Multi-algorithm

Yanhui Zhu, Ping Wang, Zhihui Wu, and ZhiQiang Wen

Institute of Computer & Communication, Hunan University of Technology,
412008 Zhuzhou, China
swayhzhu@163.com

Abstract. Semantic orientation extraction is the premise of text semantic tendency analysis and it is also the key process to obtain high precision and recall ratio. In this paper, basic semantic lexicon, conjunction corpus, and semantic distance between words are used to propose an approach of semantic orientation extraction of Chinese text based on multi-algorithm. Experiments show that the approach is better than some classical algorithms such as SO-PMI, HM, and measure of semantic distance between words.

Keywords: Multi-algorithm, Chinese text, semantic orientation, orientation extraction.

1 Introduction

Research on the theme of text classification has gained a lot of achievements for nearly a decade, but the study of the text semantic orientation is still at a relatively early stage. In recent years, many experts home and abroad have done a lot of research and developed some effective methods applying in analyzing text semantic orientation such as Bayes, maximum entropy, support vector machines and other machine learning methods. As a measure of understanding text, semantic orientation is the attitude of a particular text toward a given subject. Accurate semantic orientation and appropriate semantic weight are crucial to ensure the performance of text semantic orientation analysis, so semantic orientation extraction is one important step during the process of semantic orientation analysis.

Hatzivassiloglou and Mckeown are among those scholars who have done research on the text semantic orientation extraction and they have proposed classic HM algorithm, which identifies semantic orientation of adjectives by clustering adjectives into positive lexicon and negative lexicon on the basis of their positive or negative polarity [1]. Furthermore, it uses linguistic restrictions existing in conjunctions connecting adjectives since two adjectives connected by conjunction usually have the same or different semantic orientations. SO-PMI algorithm is proposed by Peter D. Turney which calculates the semantic orientation of a word by using the mutual information between the given word and the word "excellent" minus the mutual information between the given word and the word "poor" [2]. Some other experts like Yao Tian-Fang discovered through their research that conjunctions exist in only about

25.5 percent of valid Chinese sentences. So if HM algorithm is used in Chinese text semantic orientation extraction, we will get a poor recall ratio. In addition, our experiments show that in about 5.5 percent of Chinese sentences, adversative conjunction exists between two associations. Those facts show that the semantic words in the same sentence or neighbor semantic words may not have the same semantic orientation. That is to say, algorithms mentioned above all have their limitations [3].

To solve the problem, we use basic semantic lexicon, conjunction corpus combined with semantic distance algorithm to propose an approach of semantic orientation extraction of Chinese text based on multi-algorithm. Particularly, we extract and get preliminary Chinese semantic corpus C1 according to basic semantic lexicon first, then use conjunction corpus to extract semantic features associated with those in corpus C1 and get corpus C2. Finally, semantic features in corpus C2 are used as seed words and semantic distance algorithm is adopted to extract those semantic words with unidentified semantic orientation, which leads to form the final semantic lexicon C. The experiments show that both precision ratio and recall ratio by using our approach are better than that of classical algorithms such as SO-PMI, HM, and measure of semantic distance between words.

2 Semantic Orientation Extraction of Chinese Text Based on Multi-algorithm

2.1 Semantic Orientation Extraction Based on Basic Semantic Lexicon

Basic semantic lexicon is made of commonly used semantic words which have strong semantic orientation and are usually used as seed words for text semantic Orientation classification [4]. Due to the different expression of English text and Chinese text, we utilize ICTCLAS from Chinese academy of sciences for parsing text into words before extracting its semantic orientation [5]. Given below Fig 1. is algorithm 1 to show how to extract semantic features with the help of basic semantic lexicon.

The following instance comes from one language material about hotel comment in Tan SongBo's ChnSentiCorp - a Chinese text semantic material corpus [6]: “首先是让我满意的酒店接机服务, 这一点是和大家取得共识的了. 我是来酒钢办事的, 去酒钢办事在这里住还是非常方便的, 没有车服务员会热情周到的帮我订车. 让我比较感动的是第二天要去办事, 前一天却发现衣服挂了口子, 因为着急出去, 没时间处理, 而且本人缝纫手艺也比较差. 在情急之中找到了服务员, 回来的时候衣服补好了, 并且皮鞋也擦亮了, 一点没耽误事. 在这里对她们的服务表示小小的感谢!”

Applying algorithm 1 to this instance, we get its lexicon with semantic features C1 = (满意, 方便, 热情, 周到, 感动, 差, 耽误, 感谢). Since all these extracted words exist in basic semantic lexicon, they can express the material's semantic attitude to a certain extent.

```

// c represents basic semantic lexicon; ci represents lexicon with extracted
semantic features which initial value is NULL.
Input: document di , c, ci{ }
Output: lexicon with extracted semantic features ci{w1,w2,w3.....wk}
Begin
1. Use ICTCLAS to parse text of di into phrases
2. Preconditioning di after its parsing:
    (1) Remove all punctuation marks including the period, the question
    mark and the exclamation mark.
    (2) Remove personal names, geographic names, time as well as
    auxiliary verbs.
3. After parsing and preconditioning, to every left word wi:
    If it belongs to basic semantic lexicon, then
        we add it to ci
    End if
4. Repeat step 3 until all parsed words of di have been processed.
End
    
```

Fig. 1. Method of semantic orientation extraction based on basic semantic lexicon

2.2 Extended Semantic Features Based on Conjunction Corpus

2.2.1 Building Conjunction Corpus

In Chinese expressions, conjunctions have function as connecting words, phrases or sentences. In the case of a conjunction connecting two words, if one of them is inclined to have semantic orientation, the other word has some certain semantic tendencies too. Using this method, we can discover those unknown semantic features based on the known semantic features. At the same time, we collect conjunctions and build them into conjunction corpus.

There are only three types of conjunction categories that can be used to extract text semantic features: adversative, paratactic and progressive. These three types of conjunctions in Chinese text have great influence on the attitude of a subject but they play different role respectively. Particularly, two semantic words connected by adversative conjunction have opposite semantic orientation; However, Progressive conjunction usually connects two semantic words with the same semantic orientation, and semantic orientation of the word after conjunction is stronger than that of the word before. Similar to Progressive conjunction, paratactic conjunction connects two semantic words with the same semantic orientation. The difference between these two types of conjunction is that paratactic conjunction connecting two words of the same semantic weights. Without exception, three types of conjunction mentioned have different intensity of the semantic orientation in different context.

According to above characteristics, we use three types of conjunctive words categories: adversative, progressive and paratactic to construct conjunction corpus.

According to the characteristics of conjunctions, conjunction corpus does not tag the words value. But in order to distinguish three types of conjunctions, we assign different value to these three types conjunctions: -1 to adversative conjunction; 0 to paratactic conjunction, 1 to progressive conjunction.

Table 1. Three types of organized conjunction collections

Adversative conjunction	但, 可是, 然而, 不过, 却, 但是, 偏偏, 只是, 不过, 至于, 不料, 岂知
progressive conjunction	而且, 更加, 甚至, 不如, 不及, 乃至, 并且, 况且, 何况
paratactic conjunction	和、跟、与、同、及

2.2.2 Extend Semantic Features Based on Conjunction Corpus

Based on the basic semantic lexicon, we have already extracted semantic features from document which have strong semantic orientation, but in Chinese expression, there are other words which can express semantic orientation in addition to the basic semantic words. For example: In the sentence "your idea is good, but not feasible." , there exists a basic semantic word "good" which has the positive polarity, but you will misjudge the sentence's sentiment if you only take the basic semantic words as semantic features. The semantic orientation of this example is mainly determined by the word "not feasible". From the entire structure, the case tends to have negative semantic polarity.

```

// ci represents lexicon with extracted semantic features by using Algorithm 1;
O(w) represents semantic weight of word w
Input: document di, ci{w1,w2,w3.....wk}
Output: lexicon with extended semantic features ci{w1,w2,w3.....wm}
Begin
  1. Searching conjunctions in document di
  2. If exists conjunction conj(w):
    2.1 Searching two words w1,w2 with the same part of speech in the
    sentence including conjunction conj(w)
    2.2 If one of w1,w2 exists in ci:
      // suppose that w1 exists in ci
      (1) Add w2 to ci;
      (2) If conj(w) is paratactic conjunction:
        Assign the same emotion weight to w1,w2
      else if conj(w) is progressive conjunction:
        if w2 is before conj(w)
          O(w2) = O(w1)/2
        else O(w2) = O(w1)*2
      If conj(w) is adversative conjunction:
        O(w2) = - O(w1)
      End if
  3. Repeat step 2 until to the end of document di
End

```

Fig. 2. Extracting extended semantic features based on conjunction corpus

With conjunctions corpus, we can extract and have further semantic features based on the basic semantic features extracted by using algorithm 1.. According to the characteristics of conjunctions, we know that paratactic conjunctions connect words of the same semantic orientation with the same weight; adversative conjunctions connect words with opposite semantic orientation instead; semantic intensity of two words connected by progressive conjunction is different according to their position: the semantic intensity of the word behind progressive conjunction is stronger. Algorithm 2 as Fig 2. shows the process of extracting extended semantic features based on the conjunction corpus.

According to algorithm 2, we further extract semantic features from the text in the example of section 2.1. For example, our experiments discover a progressive conjunction "moreover" which subsequent word is an adjective "poor" existing in C1 - lexicon with basic semantic features. So we continue to search and find another semantic word "hasty" in front of the conjunction "moreover". According to preceding rules of setting semantic weight, we assign its weight the half weight of "poor". In brief, after further extraction, the extended semantic features of this instance we get are included in lexicon C2 = (满意, 方便, 热情, 周到, 感动, 差, 耽误, 感谢, 着急) .

2.3 Extend Semantic Features Further Based on Semantic Distance between Words

The basic idea of the method judging semantic distance between words is: if an article shows some semantic orientations, then the semantic words in this article will usually appear in a swarm and determine the semantic orientation of this article. In other words, the semantic orientation of most semantic words of an article is consistent. Moreover, closer semantic distance two words have, more similar semantic orientation they have [7]. In the correlation calculation of the candidate semantic words and seed words, we adopt SO-PMI algorithm (Pointwise Mutual Information).

SO-PMI algorithm evolved from PMI which is used in calculating the similarity between two elements [8]. Assume there are two words w_1 and w_2 , The Pointwise Mutual Information (PMI) between the two words is defined as formula (1), where $p(w)$ represents probability that word w exists in article and $p(w_1 \& w_2)$ represents probability that word w_1 and word w_2 coexist.

$$PMI(w_1, w_2) = \log_2 \left(\frac{p(w_1 \& w_2)}{p(w_1)p(w_2)} \right) \quad (1)$$

SO-PMI is used for judging words semantic orientation based on PMI. If there is a need to calculate semantic orientation of a given word w , then use algorithm PMI to calculate similarity between word w and every seed word respectively, after that, calculate strength of its association with k number positive words as well as with k number negative words. In other words, the semantic orientation of a given word is calculated from the strength of its association with a set of positive words, minus the strength of its association with a set of negative words. The calculation is shown as

formula (2) below: where {Kp} represents a set of words with positive semantic orientation while {Kn} represents a set of words with negative semantic orientation. If SO-PMI(w) > 0, then semantic orientation of word w is tend to be positive, otherwise, it is to be negative. Since we usually choose stronger semantic words to be seed words, bigger the absolute value of PMI (w) is, stronger the word w semantic orientation will be.

$$SO - PMI(w) = PMI(w, \{K_p\}) - PMI(w, \{K_n\}) \tag{2}$$

According to algorithm 1 and 2, we have got part of the semantic features from the text:

(1) All the basic semantic words we get by using the basic semantic lexicon are classic semantic features of text and they contribute most to the correctness of semantic orientation classification.

(2) According to the characteristics of conjunctions, we use conjunction corpus to extend the text semantic features and define their semantic weight.

Next, we use semantic distance method to extract semantic features from the rest candidate words where semantic features extracted by using algorithm 2 are taken as seed words. Specific steps are shown as follows:

① To any given candidate semantic word w in document d, we use formula (2) to calculate strength of its association with seed words and get its SO-PMI(w). If SO-PMI(w)>T (The value of T is generally set to 0.05) [7], then add w to semantic features set.

② Judge semantic orientation of w, the calculation method is shown by formula (3):

$$SO - DIST(w) = \frac{1}{\mu} \sum_i^m \log_2 \frac{N-1}{Dist(w, p_key_i)} - \frac{1}{\nu} \sum_j^k \log_2 \left(\frac{N-1}{Dist(w, n_key_j)} \right) \tag{3}$$

where $m = \overset{\circ}{\underset{\circ}{\mathbf{a}}}_i^m Dist(w, p_key_i)$ (4)

$$u = \overset{\circ}{\underset{\circ}{\mathbf{a}}}_j^k Dist(w, n_key_j) \tag{5}$$

p_key_i represents extracted positive semantic features by using algorithm 2 while n_key_j represents extracted negative semantic features. $Dist(w, p_key_i)$ represents the semantic distance between w and positive semantic features, so if it is bigger than zero, then word w is considered to have positive orientation, otherwise it tends to be negative.

③ According to the principle of semantic distance algorithm, we use the following method to calculate semantic weight of word w: If word w is considered to have

positive orientation, then we use positive semantic vocabulary of text as its weight calculation basis, otherwise we use negative vocabulary instead. Calculation method is shown by formula (6):

$$O(w) = \sum_i^k \frac{d_i}{D} * O(seed_i) \quad (6) \quad \text{where } D = \sum_i^k d_i$$

seed_i represents semantic words which have the same semantic orientation as w; O(seed_i) represents semantic weight of seed_i; d_i represents the semantic distance between w and seed_i. Algorithm is shown as follows.

After applying algorithm 3 as Fig 3. to the example in section 2.1 to further extract semantic features, we've got the following semantic features: C = (满意, 方便, 热情, 周到, 感动, 差, 耽误, 感谢, 共识, 着急, 没时间, 情急, 补好, 擦亮, 服务).

```

//ci represents semantic features set extracted by applying algorithm 2;
O(w) represents semantic weight of word w.
Input: document di; threshold t, ci{w1,w2, w3 .....wm}
Output: ci{ w1,w2,w3.....wn }
Begin
  1. Search candidate semantic word w in document di
  2. If word w does not exist in ci:
    2.1 Use formula (2) to calculate strength of its association with seed
    words.
    2.2 If SO-PMI(w)> t // t is generally set to value 0.05
      (1) Use formula (3) to judge semantic orientation of word w
      (2) Use formula (6) to calculate semantic weight of word w
      (3) Add w and its weight to ci
  3. Repeat step 1 until to the end of document di
End
    
```

Fig. 3. Judging semantic orientation based on semantic distance between words

3 Experiment and Result Analysis

The experiments use language material about hotel comments of ChnSentiCorp corpus - Chinese Sentiment Corpus from TanSongBo. The corpus include 4000 documents in total, 2000 positive and negative ones each. We selected 200 documents from the corpus randomly as training set, including 100 positive and 100 negative documents, and then tagged the semantic words manually according to their role in their context. For instance, "unhappy," is marked as negative semantic words.

The experiments identify semantic words in test corpus by utilizing semantic feature extraction method based on multi-algorithm. In addition to that, Experiments compared our semantic feature extraction method with HM algorithm, SO - PMI algorithm, semantic distance algorithms and found that the multi-algorithm made the best effect: the accuracy ratio reaches 0.940 and coverage ratio reaches 0.893. Experimental results are shown in table 2.

Table 2. Comparison between our experiment results and other semantic orientation identification methods

algorithm	accuracy ratio (%)	recall ratio (%)	F-measure (%)
HM algorithm	82.3	57.5	67.7
SO-PMI algorithm	84.6	80.4	82.5
Algorithm of semantic distance between words	83.2	72.0	77.2
Multi-algorithm	94.0	89.3	91.6

4 Conclusion

This paper proposes a method of semantic orientation extraction of Chinese text based on multi-algorithm by using basic semantic lexicon, conjunction corpus and semantic distance algorithm.

First of all, we extract basic semantic words as semantic features of text by using basic semantic lexicon, and then based on the extracted basic semantic words, we utilize the characteristics of Chinese conjunctions to extract semantic words in connection with the conjunction and take them as semantic features too; the third step is to use semantic distance algorithm to further extract semantic features in candidate semantic words with unknown semantic orientation. Experiment results show that the method is better than some classical algorithms as HM, SO - PMI, semantic distance method and etc.

Acknowledgment. This paper is sponsored by Project supported by Hunan Provincial Natural Science Foundation of China (number: 10JJ3002), China National Packaging Corporation (number: 2008-XK13) and the funds project under the Ministry of Education of the PRC for young people who are devoted to the researches of humanities and social sciences (09YJCZH019).

References

1. Hatzivassiloglou, V., McKeown, K.R.: Predicting the semantic orientation of adjectives. In: Proceedings of the 35th Annual Meeting of the ACL and the 8th Conference of the European Chapter of the ACL, pp. 174–181. Association for Computational Linguistics, Stroudsburg (1997)
2. Peter, T.: Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, pp. 417–424 (2002)
3. Tianfang, Y., Decheng, L.: Research on Semantic Orientation Distinction for Chinese Sentiment Words. In: 7th International Conference on Chinese Computing (2007)
4. Liu, W.-p., Zhu, Y.-h.: Research on building Chinese Basic Semantic Lexicon. Journal of Computer Applications 29(11), 2882–2884 (2009)

5. ICTCLAS, <http://ictclas.org/>
6. Tan, S-b.: Chinese Sentiment Corpus -ChnSentiCorp. (2010), <http://www.searchforum.org.cn/tansongbo/corpus-senti.htm>
7. Song, L., He, T.: Research on Sentiment Terms' Polarities Identification and Opinion Extraction. In: The Second Chinese Opinion Analysis Evaluation, Shanghai, pp. 30-37 (2009)
8. Li, R.-l.: Research on Text Classification and Its Related Technologies. FUDAN University, Shanghai (2005)

Formal Semantics of Chinese Discourse Based on Compositional Discourse Representation Theory

Qing-jiang Wang¹ and Lin Zhang²

¹ School of Information Engineering, North China University of Water Resources and Electric Power, Zhengzhou, China
wangqingjiang@ncwu.edu.cn

² The Modern Education Technology Center, Henan University of Economics and Law, Zhengzhou, China
zhanglin@hnufe.edu.cn

Abstract. To obtain formal semantics of Chinese discourse, compositional discourse representation theory was adopted to compose semantics from Chinese word to sentence and to discourse. According to syntactic rules and λ -box expressions of word categories, sentential logic expression could be created by translating from child nodes to father node in syntactic tree, then the formal representation of Chinese discourse could be further obtained by using construction procedure of DRT to resolve cross-sentential demonstrates and merge sentential DRs. A rough prototype system was given to depict the implementation process, which will benefit establishing machine-inside representation of discourse semantics of contemporary Chinese.

Keywords: compositional discourse representation theory, Chinese discourse, formal semantics.

1 Introduction

The Internet has been the dominant discourse platform for billions of Chinese people, and the large amount of written-language information urgently needs understanding-based processes by computers rather than by human reading, to implement efficient network information supervision, network public sentiment investigation, network politics consultant, etc.

The level computers understand natural language depends upon research progress of many disciplines including modern linguistics, linguistic philosophy, linguistic logic, artificial intelligence, and cognitive science, especially in their aspects about formal syntax, formal semantics, and formal pragmatics. The three division of syntax, semantics and pragmatics clarifies aim-specific three layers from intricate linguistic phenomena. The syntax layer focuses on identifying word categories and analysing sentence structure, the semantics layer emphasizes sentence meaning representation, and the pragmatics layer answers for resolving cross-sentential demonstrates, representing discourse meaning, and implementing semantic reasoning, which reflects human thinking processes and final discourse understanding.

This paper reviews classical achievements about formal syntax, formal semantics and formal pragmatics, and discusses how to utilize compositional discourse representation theory to obtain formal semantics of Chinese discourse, from which logic reasoning may be further fulfilled.

2 Related Works

The ambiguity and variety made natural languages not be thought as a formal analysis-enable system until N. Chomsky presented his syntactic creative and transforming theory in 1950s, which explains the relationship between infinite structure-style sentences and finite syntactic rules. To 1990s, Chomsky and his associates successively presented Standard Theory, Extended Standard Theory, Government and Binding Theory, and Minimalist Program, forming relatively perfect theory system of syntactic creation and transformation [1]. Chomsky emphasizes syntax investigation, and seldom deals with semantics and pragmatics.

R. Montague proposed Universal Grammar in 1970s, thereafter called Montague Grammar, and argued that there is no important theoretical differences between natural languages and the artificial languages of logicians, and it is possible to comprehend the syntax and semantics of both kinds of language within a single natural and mathematically precise theory [2]. In Montague Grammar, categorial operations correspond with syntactic concatenation, the rules from syntactic categories to intensional logic are designed for translating English sentences to type logic expressions, and model-theoretical semantics is used to explain the meaning of English sentences. Syntactic concatenation corresponds with categorial operations, and the latter corresponds with λ -transformation in type logic, so type logic expressions of English sentences are obtained when sentences are created from bottom to top, the process of which implements semantic combination of syntactic components, namely, the meaning of a complex syntactic structure is a function of the meanings of its parts.

Sentence semantics are handled statically in Montague Grammar, and anything as the relationship between a pronoun and its antecedent is no longer attached to type expressions once they are translated from sentences. Montague Grammar pays attention to semantic composition under sentence level, namely how to compose sentence meaning from word meaning, neglecting semantic relations across sentences, consequently it can't resolve demonstrates in discourse, such as donkey sentence problems.

At 80s, H. Kamp depicted Discourse Representation Structure (DRS), a medium layer between syntax layer and its semantic model, to represent semantic results obtained by dynamic and gradual analyses, which is called Discourse Representation Theory (DRT) [3][4][5]. DRT consists of two parts, one is recursive definition of DRS and its model-theoretic semantics, and the other is construction procedure which specifies how to extend a given DRS when a sentence comes in. DRSs are set-theoretic objects built from discourse referents and DRS-conditions, and may be obtained in two steps. In the first one, a sentence DRS was constructed in a compositional way, and in the second step, the sentence DRS was merged with the DRS representing the prior discourse, and anaphoric references were resolved. Apparently, two-stage construction procedure reflects the division between semantics and

pragmatics. DRT is the first discourse semantic theory, which can explain cross-sentential pronominal anaphora, tense, presupposition, etc.

Traditional DRT specifically deontes Kamp's DRT and its developments, and generalized DRT includes File Change Semantics (FCS) and Dynamic Semantics (DS). At 1982, Irene Heim proposed FCS to dynamically analyze context-dependend discourse semantics by introducing context into Montague semantics, which broke through the boundedness of static meaning analyses in traditional formal semantics, and could explain the references of definite and indefinite noun phrases [6].

At 1989, J. Groenendijk and M. Stokhof found there was a tendency to do away with the compositional principle in various discourse semantics theories including DRT, while compositionality was not only a methodological question, but also could explain more linguistic phenomenons, thus compositional discourse semantics theory should be proposed to substitute those non-compositional ones, and that is Dynamic Predicate Logic (DPL) [7]. DPL remains syntax of predicate logic, but adopts dynamic semantics which can explain cross-sentential pronominal anaphora.

At 1996, R. Muskens presented to reduce DRS into type logic, introduce λ -operator and adopt dynamic semantics with the purpose of implementing semantic composition from word to sentence and to discourse, that is Compositional DRT (CDRT) [8]. The basic types in CDRT type logic include entity type e , truth value type t , register type π , and state type s . Registers store discourse referents, resembling variables in program language. A state may be thought of as a list of the current inhabitants of all registers, so states are very much like the program states that the theoretical computer scientists talk about. $i[x_1x_2\dots x_n]j$ expresses that i and j differ at most in x_1, x_2, \dots , and x_n , $V(x)(i)$ denotes the value of x at state i , and state i may be represented as $\lambda x(V(x)(i))$. According to (1), DRS syntax can be reduced to type logic. By assigning type to each word category and determining their type logic expressions, semantics composition may be implemented from words to discourse.

$$\begin{aligned}
 R(x_1, \dots, x_n) &\mapsto \lambda i. R(V(x_1)(i), \dots, V(x_n)(i)) \\
 x_1 \text{ is } x_2 &\mapsto \lambda i. V(x_1)(i) = V(x_2)(i) \\
 \text{not } K &\mapsto \lambda i. \neg \exists j K(i)(j) \\
 K \text{ or } K' &\mapsto \lambda i \exists j (K(i)(j) \vee K'(i)(j)) \\
 K \Rightarrow K' &\mapsto \lambda i \forall j (K(i)(j) \vee \exists k K'(j)(k)) \\
 [x_1 \dots x_n | \gamma_1 \dots \gamma_m] &\mapsto \lambda i \lambda j (i[x_1 \dots x_n]j \wedge \gamma_1(j) \wedge \dots \wedge \gamma_m(j)) \\
 K ; K' &\mapsto \lambda i \lambda j \exists k (K(i)(k) \wedge K'(k)(j))
 \end{aligned} \tag{1}$$

Both static semantics in Montague grammar and dynamic semantics in CDRT are attributed to model-theoretical semantics, which considers word meaning as atomically inseparable. Nouns are explained as individual sets, transitive verbs are explained as the set of individual-ordered pairs, and words with other categories are also defined by themselves.

The institute of computational linguistics in Peking University researched Chinese information processing during recent 20 years, established grammatical knowledge-base of contemporary Chinese including 73,000 entries, semantic knowledge-base of

contemporary Chinese including 66,000 entries, Chinese concept dictionary, and shifted research emphases from syntax to semantics and pragmatics. Ping Wu analyzed formal semantics of some Chinese sentences using formal linguistics [9], and Xiang-hong Fang investigated the semantics of Chinese connectors and correlative constructions based on intensional logic [10], Chong-li Zou constructed sentence system about Chinese body features based Montague grammar [1].

From the above, homomorphic operations of syntax and semantics in Montague grammar has far-reaching significance. Together with Chomsky’s or Montague’s creative syntax, semantics shows compositionality and calculability. DRT dominates discourse semantics representation, and can implement semantic composition from word to sentence and to discourse after introducing type logic and adopting dynamic semantics. Most Chinese semantics remained at formal semantic analyses using Montague grammar, and haven’t investigated further discourse semantics with DRT.

3 Transforming Chinese Discourse to Its Formal Semantics

Contemporary Chinese has some different linguistic phenomena with English, but Chinese also has syntactic creativity, although systematic creative and transforming syntax for Chinese has not been found yet. Syntactic creativity makes it possible of transforming Chinese discourse to formal semantics by CDRT.

It’s feasible and promising of designing syntactic creative rules by investigating relationship between categorial operations and syntactic concatenations. Once creative and transforming rules for Chinese syntax being established, Chinese words could be assigned λ -box expressions of DRT according to word categories and syntactic rules, illustrated as (2).

$$\begin{aligned}
 \text{杨利伟}_n &\mapsto \lambda P.P(\text{杨利伟}) \\
 \text{他}_n &\mapsto \lambda P.P(\delta), \text{ here } \delta = dr(\text{ant}(\text{他}_n)) \\
 \text{是} &\mapsto \lambda P\lambda v[|P(v)] \\
 \text{中国人的} &\mapsto \lambda P\lambda v[| \text{中国人的}(v)]; P(v) \\
 \text{梦想} &\mapsto \lambda v[| \text{梦想}(v)] \\
 \text{实现} &\mapsto \lambda Q\lambda v(Q(\lambda v'[| \text{实现}(v, v')]))
 \end{aligned}
 \tag{2}$$

Here superscript n denotes the n th discourse referents, and subscript n denotes the n th pronoun. Suppose $dr(A_n)$ returns discourse referent u_n introduced by indefinite or definite noun A_n , $\text{ant}(B_n)$ returns discourse referent A_n denoted by pronoun B_n , then $dr(\text{ant}(B_n))=u_n$. For example, $dr(\text{ant}(\text{他}_1))=\text{杨利伟}^1$. The relationship between discourse referents and pronouns is determined by construction procedure of DRT adapted for Chinese syntax.

$$\text{杨利伟}^1 \text{实现了中国人的飞天梦想, 他}_1 \text{是英雄.} \tag{3}$$

According to syntactic trees, there are following rules and merging lemma for accomplishing translation from child nodes to father nodes. The translation process of (3) is illustrated as Fig. 1, and the root node is the final result.

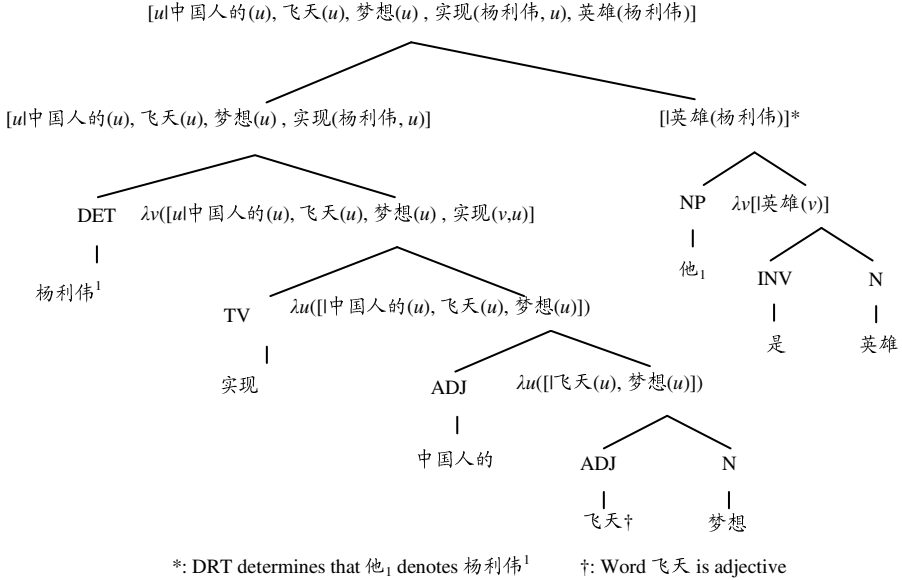


Fig. 1. Compositionally create discourse semantics of contemporary Chinese

R1(replication rule): if $A \mapsto \alpha$ and A is the only child node of B , then $B \mapsto \alpha$

R2(transformation rule): if $A \mapsto \alpha$, $B \mapsto \beta$, and A and B are only child nodes of C , then $C \mapsto \alpha(\beta)$

R3(sequencing rule): if $A \mapsto \alpha$, $B \mapsto \beta$, and A and B are child nodes of C , then $C \mapsto \alpha;\beta$

R4(quantifying rule): if $NP_n \mapsto \alpha$, $B \mapsto \beta$, and NP and B are child nodes of C , then $C \mapsto \alpha(\lambda v_n \beta)$

R5(reduction rule): if $A \mapsto \alpha$, and β follows from α by reduction, then $A \mapsto \beta$

Merging lemma: if x_1, x_2, \dots, x_k do not occur in any of $\gamma_1, \gamma_2, \dots, \gamma_m$, then

$$\| [y_1, \dots, y_n \mid \gamma_1, \dots, \gamma_m] \|; \| [x_1, \dots, x_k \mid \delta_1, \dots, \delta_q] \| = \| [y_1, \dots, y_n, x_1, \dots, x_k \mid \gamma_1, \dots, \gamma_m, \delta_1, \dots, \delta_q] \|$$

From (1), the root node in Fig. 1 has type logic expression as (4).

$$\lambda i \lambda j (i[u]j \wedge \text{中国人的}(V(u)(j)) \wedge \text{飞天}(V(u)(j)) \wedge \text{梦想}(V(u)(j)) \wedge \text{实现}(V(\text{杨利伟})(j), V(u)(j)) \wedge \text{英雄}(V(\text{杨利伟})(j))) \quad (4)$$

According the truth definition of DRS and unselective binding lemma in CDRT, Formula (4) can be transformed to (5), from which semantic reasoning could be conducted by deductive system of first order logic.

$$\exists x (\text{中国人的}(x) \wedge \text{飞天}(x) \wedge \text{梦想}(x) \wedge \text{实现}(\text{杨利伟}, x) \wedge \text{英雄}(\text{杨利伟})) \quad (5)$$

A rough prototype system for investigating Chinese discourse semantics is as Fig. 2. According to syntax struction trees of Chinese discourse and λ -box expressions of Chinese words, semantics-compositional module creates formal representation of discourse semantics, which could be further transformed to first order logic.

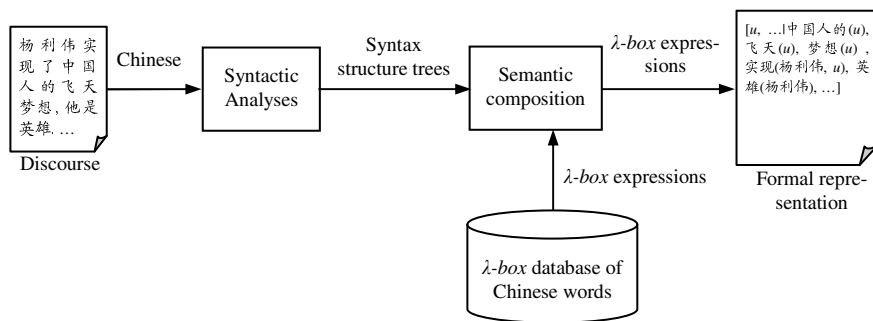


Fig. 2. Prototype system for transforming Chinese discourse to its formal semantics

4 Conclusions

Chomsky argued linguistic abilities of human being is inborn and from evolution, and syntax includes deep structures and surface structures. Deep structures are created from phrase concatenations according to some syntactic rules, and surface structures are transformed from deep structures. Here considered that creative and transforming syntax is also existing for contemporary Chinese, although such syntactic system has not been established yet. Besides, other achievements of linguistic philosophy including CDRT could also benefit machine understanding of Chinese, and here gives a rough prototype system to get formal semantics of Chinese discourse, which will be helpful for establishing machine-inside representation of discourse semantics of contemporary Chinese and constructing world knowledge base from large amounts of Chinese discourses.

References

1. Cai, S.-s., Zou, C.-l.: Study on Formal Theories of Natural Languages. People Pub., Beijing (2010) (in Chinese)
2. Montague, R.: Formal Philosophy: Selected Papers. Yale University Press, New Haven (1974)
3. Eijck, J.: Discourse Representation Theory, <http://homepages.cwi.nl/~jve/papers/05/drt/drt.pdf>
4. Bart, G., David, I.B.: Discourse Representation Theory, <http://plato.stanford.edu/entries/discourse-representation-theory/>
5. Kamp, H., Reyle, U.: A Calculus for First Order Discourse Representation Structures. Journal of Logic, Language, and Information 5, 297–348 (1996)
6. Heim, I.: The Semantics of Define and Indefinite Noun Phrases. Garland Pub., New York (1988)
7. Groenendijk, J., Stokhof, M.: Dynamic Predicate Logic. Linguistics and Philosophy 14, 39–100 (1991)

8. Muskens, R.: Combining Montague Semantics and Discourse Representation. *Linguistics and Philosophy* 19, 143–186 (1996)
9. Wu, P.: Formal Semantic Analysis of Selected Constructions in Mandarin Chinese. Dissertation of Beijing Language and Culture University (2005) (in Chinese)
10. Fang, X-h.: An Intensional Logic Based Semantic Study on Chinese Connectors and Correlative Constructions. Dissertation of Shanghai Normal University (2004) (in Chinese)

Research of Chinese Word Sense Disambiguation Based on HowNet

Jingwen Zhan and Yanmin Chen

College of Information Engineering, China Jiliang University, Hangzhou, 310018, China
cissy331@msn.com, chenyanmin@gmail.com

Abstract. A Chinese word sense disambiguation algorithm based on HowNet is proposed in this paper. After extracting the ambiguous words in the text, the correlative words which restrict the word sense in the context can be found by using dependency grammar analysis. Finally, by computing the semantic relevancy between the atomic term of ambiguous words and the correlative words, the correct meaning of the ambiguous word can be determined by the value of the semantic relevancy. Experimental results show that the method has a good effect on Chinese word sense disambiguation.

Keywords: Word Sense Disambiguation, Natural Language Processing, HowNet, Dependency Grammar Analysis, Semantic Relevancy.

1 Introduction

There are a large number of multi-meaning words in natural language, and how to determine the correct word meaning automatically according to a given context is just the problem which should be solved in word sense disambiguation. Word sense disambiguation has always been a hot research issue of natural language processing. As an indispensable link in natural language processing, word sense disambiguation is widely used in many fields. It plays a very important role in the fields of machine translation, information retrieval, text analysis, automatic summarization, knowledge mining, and so on [1]. As the widespread distribution of polysemy in natural language, word sense disambiguation is considered as one of the most difficult problems in natural language processing. Since the sparse data in knowledge acquisition and the parameter space of automatic learning algorithm is too big, so far, there are few methods which can solve the problem of word sense disambiguation completely [2].

In this paper, an algorithm of Chinese word sense disambiguation based on HowNet is proposed. After extracting the ambiguous words in the text, in order to find the correlative words, which restrict the word sense, the fully dependency grammar analysis is adopted to find dominant and dominated relation among words from inner structure of the context. By computing the semantic relevancy between the atomic term of ambiguous words and the correlative words, the correct meaning of the ambiguous word would be determined by the value of the semantic relevancy. And finally, we output the disambiguated text as the final result.

2 HowNet

HowNet is an on-line common-sense knowledge base unveiling inter-conceptual relations and inter-attribute relations of concepts as connoting in lexicons of the Chinese and their English equivalents [3]. It not only marked the semantics and the part of speech of each word, the concept which the word belongs to, the links between concept and its internal attributes and the links between the concepts are marked as well. Now, HowNet has become one of the most important resources of natural language processing.

There are two main points in HowNet: concept and sememe. Concept is a description of lexical semantics. Multi-meaning word can be expressed by several different concepts. And sememe is the minimum unit which is used to describe a concept [4]. As a knowledge base, the knowledge structured by HowNet is a graph rather than a tree. It is devoted to demonstrate the general and specific properties of concepts.

3 Dependency Grammar Analysis

Describing the structure of language by dependencies between words, such a framework is called dependency grammar, which is called affiliation too [5]. At present, dependency grammar is one of the hot topics in the research of syntax. The so-called *dependency* means dominant and dominated relation between words. Dependency grammar reveals the syntactic structure by analyzing the dependencies between the components in the language unit. Through dependency grammar analysis, those words which have dominant and dominated relation with the ambiguous words in the context can be extracted, and these words are defined correlative words of the ambiguous words [6].

The extracting of correlative words in the context has a directly impact on the result of disambiguation. Compared to the traditional “context window” approach[7], getting correlative words by dependency grammar analysis can reduce the noise which made by the irrelevant context, decrease the computational times significantly, improve the accuracy of disambiguation and the efficiency at the same time.

4 Computing of Semantic Relevancy

Since the disambiguation in this paper is taken by analyzing the ambiguous words and their correlative words, the semantic relevancy between the atomic term of ambiguous words and the correlative words has become a key indicator to the final decision. As we known, the main influencing factors of semantic relevancy including similarity, relevance and the factor of examples [8]. In this paper, the compute of semantic relevancy is taken between the atomic term of ambiguous words and the correlative words.

4.1 The Concept Similarity

In HowNet, all the concepts are described by sememe, so the similarity of sememe is the basis. The similarity between two sememes [9] can be calculated as follow.

$$Sim(p_1, p_2) = \alpha / (d + \alpha) \tag{1}$$

Where, d is the path distance between the sememe p_1 and p_2 . α is the value of path distance when the semantic distance between two sememes is 0.5.

Since each concept of a word can be expressed by several sememes, and these sememes can be classified into four categories which can be marked as $Sim_1(S_1, S_2)$, $Sim_2(S_1, S_2)$, $Sim_3(S_1, S_2)$, $Sim_4(S_1, S_2)$, the similarity of concept can be computed as follow.

$$Sim(S_1, S_2) = \sum_{i=1}^4 \beta_i \prod_{j=1}^i Sim_j(S_1, S_2) \tag{2}$$

Where, $\beta_1 + \beta_2 + \beta_3 + \beta_4 = 1, \beta_1 \geq \beta_2 \geq \beta_3 \geq \beta_4$.

4.2 The Concept Relevance

According to the structure of HowNet, the relevance between sememes is formed by the horizontal relationship between a sememe and its explanation sememe. And it can be computed as follow.

$$rele(p_1, p_2) = \sum_{p_1 \in \exp(p_1)} \frac{W_i * Sim(p_i, p_2)}{n} + \sum_{p_2 \in \exp(p_2)} \frac{W_j * Sim(p_1, p_j)}{m} \tag{3}$$

Where, $\exp(p_1)$ and $\exp(p_2)$ are sets of explanation sememes for sememe p_1 and p_2 . W is the weight of the relationship between two sememes. n and m are the number of explanation sememe of p_1 and p_2 respectively. On this basis, the relevance of concept can be computed as follow.

$$Rele(S_1, S_2) = MaxRele(p_i, p_j) \tag{4}$$

4.3 The Factor of Examples

In HowNet, a number of common words have Chinese examples. They are useful for us to understand the meaning more vividly. And they also have impact on the compute of relevancy. In this paper, the factor of examples is marked as E , which can be computed as follow.

$$E(S_1, S_2) = maxSim(Se_i, S_j) \quad 1 \leq i, j \leq 2, i \neq j \tag{5}$$

An ambiguous word has several meaning or concept, each concept has its own examples and each example includes several words. Se_i is a concept of an example word.

According to all of these, the formula of compute semantic relevancy can be expressed as follow.

$$R(S_1, S_2) = \gamma_1 * Sim(S_1, S_2) + \gamma_2 * Rele(S_1, S_2) + \gamma_3 * E(S_1, S_2) \tag{6}$$

Where, γ_1, γ_2 , and γ_3 are the different weight for each part in this formula and $\gamma_1 + \gamma_2 + \gamma_3 = 1$.

5 The Process of Disambiguation System

The process of Chinese word sense disambiguation based on HowNet can be expressed by the flow chart as follow.

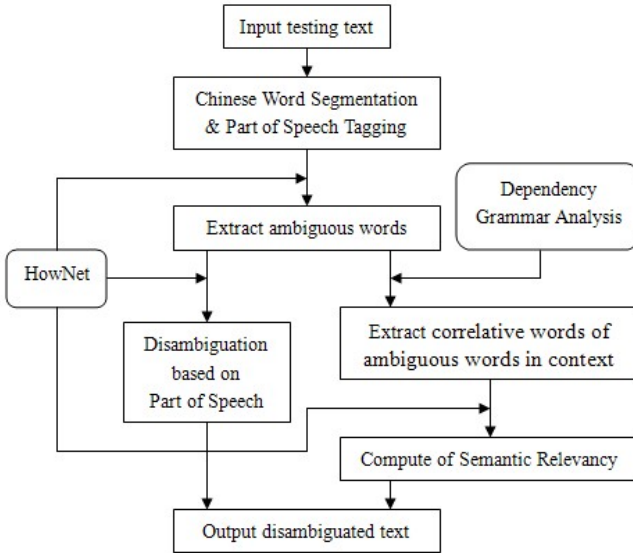


Fig. 1. The flow chart of the WSD system

In this disambiguation system, firstly, took a text preprocessing to the testing text, including Chinese word segmentation and part of speech tagging. And then combined it with HowNet to extract the ambiguous words, which have more than one meanings. These ambiguous words are to be disambiguated.

For these ambiguous words, if the ambiguous word just has one meaning under the part of speech which is tagged in the text, it can directly output the meaning as the disambiguated result. Other ambiguous words will be further disambiguated by the context, getting correlative words by dependency grammar analysis, computing the semantic relevancy between the atomic term of ambiguous words and the correlative words by (6), taking the atomic term of ambiguous word which gets the maximum semantic relevancy as the correct one, and finally outputting the disambiguated text.

From the flow chart, it shows that blocks of text preprocessing, dependency grammar analysis and compute of semantic relevancy are all affective to the final disambiguation result directly.

6 Experiment and Evaluation

Because the disambiguation in this paper is for Chinese words, 5 common Chinese words are chosen to take the test of disambiguation. The result of our method is to

compare to the algorithm based on word similarity for evaluation. We extract the five common words in a number of documents, and input them into both our method and the algorithm based on word similarity. Using the results from both methods, a precision can be calculated as follow.

$$precision = AW / DW \quad (7)$$

Where, AW means the number of ambiguous words which are disambiguated correctly, DW means the number of all ambiguous words in the testing text.

Table 1 shows the results comparing our method with the method based on word similarity [10].

Table 1. Comparing the results of our method and the method based on word similarity

Ambiguous words	Precision	
	Our method	Method based on word similarity
“辉煌”	0.68	0.59
“分子”	0.66	0.52
“材料”	0.61	0.37
“穿”	0.69	0.41
“运动”	0.62	0.46
AVG	0.61	0.47

The results of Table 1 show that the precision of our method is higher. That means our method has a good effect on Chinese word sense disambiguation.

7 Conclusion

In this paper, an algorithm of Chinese word sense disambiguation based on HowNet is proposed. Getting correlative words of ambiguous words in the context by dependency grammar analysis, calculating the semantic relevancy between the atomic term of ambiguous words and the correlative words based on HowNet, and deciding the correct meaning of the ambiguous word according to the value of semantic relevancy. By introducing relevancy, the paper uses factor of examples and the relationship between horizontal and vertical directions of HowNet semantic knowledge resource to compute the relevancies of different words, so as to improve the accuracy of relevancy. Revealed from the analysis of the experimental data, this method is efficient on Chinese word sense disambiguation.

References

1. Schutze, H., Pedersen, J.: Information Retrieval Based on Word Senses. [EB/OL] (2007)
2. Lan, M.-h., Li, H.-l.: Methods of Word Sense Disambiguation. Science & Technology Information 9, 478–482 (2010)

3. Dong, Z.-d., Dong, Q.: HowNet, <http://www.keenage.com>
4. Liu, Q., Li, S.: Word Similarity Computing Based on How-net. In: The Third Chinese Lexical Semantics Workshop, Taipei (2002)
5. Yuan, W.: A Summary of Dependency Grammar. *SCI-TECH Information Development & Economy* 20, 152–154 (2010)
6. Jing, S., Li, W.-l.: Research on Approach of Chinese Semantic Analysis. *Application Research of Computers* 27, 529–531 (2010)
7. Zhan, J., Chen, Y.: Research on Word Sense Disambiguation. In: International Conference on Materials Science & Technology, Jeju island, Korea, pp. 337–342 (2010)
8. Shengqi, L., Qiaoyan, T., Cheng, T.: Disambiguating Method for Computing Relevancy based on HowNet Semantic Knowledge. *Journal of the China Society for Scientific and Technical Information* 28, 706–711 (2009)
9. Peng, L., Sun, M.-y.: HowNet-Based Word Sense Automatic Classification System. *Computer Simulation* 2, 127–131 (2004)
10. Yu, X., Liu, P., Zhao, T.: A Method of Chinese Word Sense Disambiguation Based on HowNet. In: The Second Session of the Student Computer Linguistics Conference, Beijing, pp. 128–133 (2004)

Fast Quantum Algorithm of Solving the Protein Folding Problem in the Two-Dimensional Hydrophobic–Hydrophilic Model on a Quantum Computer

Weng-Long Chang

Department of Computer Science and Information Engineering, National Kaohsiung University of Applied Sciences, 415 Chien Kung Road, Kaohsiung 807, Taiwan, R.O.C.

changwl@cc.kuas.edu.tw

Abstract. In this paper, it is first demonstrated that for given a linear chain of amino acids the quantum Boolean circuit of implementing the function of its folding in the two-dimensional hydrophobic–hydrophilic model is responsible for processing all the computational basis states and labeling the unique answer (its unique native structure). Next, it is shown that amplitude amplification of the unique answer can be completed by means of Shor’s quantum order-finding algorithm. Then, it is also proved that after a measurement on the unique answer is completed, the successful probability of obtaining the unique answer is the same as that of Shor’s quantum order-finding algorithm.

Keywords: Protein Folding, Quantum Algorithms, Quantum Circuits, Quantum Computing, NP-Complete Problems.

1 Quantum Algorithms of Solving the Protein Folding Problem in the Two-Dimensional Hydrophobic-Hydrophilic Model

1.1 Construction of Conformational Space to a Protein in the Hydrophobic-Hydrophilic Model in the Two-Dimensional Lattice

In the hydrophobic–hydrophilic model in a two-dimensional lattice, for generating all of the possible conformations for a protein with an amino acid sequence $\mathbf{a} = a_1 \dots a_L$, $(2 \times (L - 1))$ Hadamard gates are used to operate the $(2 \times (L - 1))$ quantum bits, $(\otimes_{o=L}^2 (\otimes_{p=2}^1 |w_{o,p}^0\rangle))$, in the quantum state vector, $(|\Omega_0\rangle)$, and the following new quantum state vector, $(|\Omega_{6-1}\rangle)$, is obtained.

$$\begin{aligned}
 |\Omega_{6-1}\rangle &= (\otimes_{e=L}^2 (\otimes_{q=\theta}^1 |d_{e,q}^0\rangle)) \otimes (\otimes_{q=\theta}^1 |b_{e,q}^0\rangle) \otimes \left(\frac{1}{\sqrt{2^{2 \times (L-1)}}}\right) \\
 &\left(\sum_{w=0}^{2^{2 \times (L-1)} - 1} (\otimes_{e=1}^1 (\otimes_{q=\theta}^1 |d_{e,q}^0\rangle) \otimes (\otimes_{q=\theta}^1 |b_{e,q}^0\rangle)) \otimes (\otimes_{q=\theta}^2 |\alpha_q^0\rangle) \otimes (\otimes_{q=1}^1 |\alpha_q^1\rangle)\right) \\
 &\otimes (\otimes_{p=\theta}^0 |c_p^0\rangle) \otimes (|w\rangle). \tag{1-1}
 \end{aligned}$$

This indicates that in (1-1) in the new quantum state vector, $|\Omega_{6-1}\rangle$, the position of the first lattice site in $(2^{2 \times (L-1)})$ possible conformations is $(\otimes_{e=1}^1 (\otimes_{q=\theta}^1 |d_{e,q}^0\rangle)) \otimes (\otimes_{q=\theta}^1 |b_{e,q}^0\rangle)$, where $(\otimes_{e=1}^1 (\otimes_{q=\theta}^1 |d_{e,q}^0\rangle))$ is used to encode the value of the horizontal coordinate and $(\otimes_{e=1}^1 (\otimes_{q=\theta}^1 |b_{e,q}^0\rangle))$ is applied to encode the value of the vertical coordinate for the first lattice site in a two-dimensional lattice. For the convenience of our presentation, it is assumed that for $1 \leq e \leq L$ and $1 \leq q \leq \theta$, $|D_e\rangle$ is applied to represent the corresponding decimal value of $|d_{e,\theta} \dots d_{e,1}\rangle$ and $|B_e\rangle$ is used to represent the corresponding decimal value of $|b_{e,\theta} \dots b_{e,1}\rangle$.

It is assumed that $(|D_o\rangle, |B_o\rangle)$ and $(|D_{o-1}\rangle, |B_{o-1}\rangle)$ are, respectively, the two-dimensional coordinate of the $(o-1)^{th}$ lattice site and the two-dimensional coordinate of the o^{th} lattice site in a two-dimensional lattice. From the $(o-1)^{th}$ lattice site to the o^{th} lattice site in a two-dimensional lattice, four two-dimensional coordinates, $(|D_o\rangle, |B_o\rangle)$, to the up movement, the right movement, the down movement and the left movement are, respectively, $(|D_{o-1}\rangle, |B_{o-1}+1\rangle)$, $(|D_{o-1}+1\rangle, |B_{o-1}\rangle)$, $(|D_{o-1}\rangle, |B_{o-1}-1\rangle)$ and $(|D_{o-1}-1\rangle, |B_{o-1}\rangle)$. A unitary gate, U_{CS} , is applied to finish computation of the corresponding coordinates of four movements from the $(o-1)^{th}$ lattice site to the o^{th} lattice site in a two-dimensional lattice. The following algorithm is used to construct conformational space of a protein with an amino acid sequence $\mathbf{a} = a_1 \dots a_L$ in the hydrophobic-hydrophilic model in a two-dimensional lattice.

Algorithm 1-1: Construct conformational space of a protein with an amino acid sequence $\mathbf{a} = a_1 \dots a_L$ in the hydrophobic-hydrophilic model in a two-dimensional lattice.

(1) For $o = 2$ to L

(1a) Based on that $(\otimes_{q=\theta}^1 |d_{o,q}\rangle)$ and $(\otimes_{q=\theta}^1 |b_{o,q}\rangle)$ are target bits and $(\otimes_{q=\theta}^1 |d_{o-1,q}\rangle)$ and $(\otimes_{q=\theta}^1 |b_{o-1,q}\rangle)$ are control bits, a unitary gate, $(\otimes_{q=\theta}^1 |d_{o,q}^0 \oplus d_{o-1,q}\rangle) \otimes (\otimes_{q=\theta}^1 |b_{o,q}^0 \oplus b_{o-1,q}\rangle)$, is used to operate the quantum state vector $|\Omega_{6-(o-1)}\rangle$, and the following new quantum state vector is obtained:

$$\begin{aligned}
 |\Omega_{6-o,1}\rangle &= (\otimes_{e=L}^{o+1} (\otimes_{q=\theta}^1 |d_{e,q}^0\rangle)) \otimes (\otimes_{q=\theta}^1 |b_{e,q}^0\rangle) \otimes \\
 & \left(\frac{1}{\sqrt{2^{2 \times (L-1)}}} \left(\sum_{w=0}^{2^{2 \times (L-1)} - 1} (\otimes_{e=o}^o (\otimes_{q=\theta}^1 |d_{e,q}^0 \oplus d_{e-1,q}\rangle) \otimes (\otimes_{q=\theta}^1 |b_{e,q}^0 \oplus b_{e-1,q}\rangle) \right) \right) \\
 & \otimes (\otimes_{e=o-1}^1 (\otimes_{q=\theta}^1 |d_{e,q}\rangle) \otimes (\otimes_{q=\theta}^1 |b_{e,q}\rangle)) \otimes (\otimes_{q=\theta}^2 |\alpha_q^0\rangle) \otimes (\otimes_{q=1}^1 |\alpha_q^1\rangle) \otimes \\
 & (\otimes_{p=\theta}^0 |c_p^0\rangle) \otimes (|w\rangle).
 \end{aligned}$$

(1b) A unitary gate, U_{CS} , is applied to figure out the corresponding coordinates of four movements from the $(o - 1)^{th}$ lattice site to the o^{th} lattice site in a two-dimensional lattice and is used to operate the quantum state vector $|\Omega_{6-o,1}\rangle$, and the following new quantum state vector is obtained:

$$\begin{aligned}
 |\Omega_{6-o}\rangle &= (\otimes_{e=L}^{o+1} (\otimes_{q=\theta}^1 |d_{e,q}^0\rangle)) \otimes (\otimes_{q=\theta}^1 |b_{e,q}^0\rangle) \otimes \left(\frac{1}{\sqrt{2^{2 \times (L-1)}}}\right) \\
 &\left(\sum_{w=0}^{2^{2 \times (L-1)} - 1} (\otimes_{e=o}^o (\otimes_{q=\theta}^1 |d_{e,q}\rangle) \otimes (\otimes_{q=\theta}^1 |b_{e,q}\rangle)) \otimes (\otimes_{e=o-1}^1 (\otimes_{q=\theta}^1 |d_{e,q}\rangle) \otimes \right. \\
 &\left. (\otimes_{q=\theta}^1 |b_{e,q}\rangle)) \otimes (\otimes_{q=\theta}^2 |\alpha_q^0\rangle) \otimes (\otimes_{q=1}^1 |\alpha_q^1\rangle) \otimes (\otimes_{p=\theta}^0 |c_p^0\rangle) \otimes (|w\rangle)\right).
 \end{aligned}$$

EndFor
EndAlgorithm

Lemma 1-1: Algorithm 1-1 can be applied to construct conformational space of a protein with an amino acid sequence $\mathbf{a} = a_1 \dots a_L$ in the hydrophobic–hydrophilic model in a two-dimensional lattice.

1.2 Computation of the Global Energy for Each Possible Conformation from Conformational Space to a Protein in the Hydrophobic-Hydrophilic Model in the Two-Dimensional Lattice

It is assumed that $(4 \times L \times L)$ auxiliary quantum registers of one quantum bit, $|\Phi_{k,j,p}\rangle$, for $1 \leq k \leq L$, $1 \leq j \leq L$ and $0 \leq p \leq 3$ are used to record whether the two-dimensional coordinate of the k^{th} lattice site is adjacent to the two-dimensional coordinate of the j^{th} lattice site. Also, it is supposed that $|\Phi_{k,j,p}^1\rangle$ denotes the fact that the value of $|\Phi_{k,j,p}\rangle$ is 1, and $|\Phi_{k,j,p}^0\rangle$ denotes the fact that the value of $|\Phi_{k,j,p}\rangle$ is 0. Also it is assumed that $|\Phi_{k,j,0}^1\rangle$, $|\Phi_{k,j,1}^1\rangle$, $|\Phi_{k,j,2}^1\rangle$ and $|\Phi_{k,j,3}^1\rangle$ denote the up neighbor, right neighbor, down neighbor, and left neighbor, respectively, for the k^{th} lattice site and the j^{th} lattice site. The initial state for $|\Phi_{k,j,p}\rangle$, for $1 \leq k \leq L$, $1 \leq j \leq L$ and $0 \leq p \leq 3$ is set to $|\Phi_{k,j,p}^0\rangle$.

After Algorithm 1-1 is completed, the following quantum state vector is obtained:

$$\begin{aligned}
 |\Omega_{6-L}\rangle &= \left(\frac{1}{\sqrt{2^{2 \times (L-1)}}}\right) \left(\sum_{w=0}^{2^{2 \times (L-1)} - 1} (\otimes_{e=L}^1 (\otimes_{q=\theta}^1 |d_{e,q}\rangle) \otimes (\otimes_{q=\theta}^1 |b_{e,q}\rangle)) \otimes \right. \\
 &\left. (\otimes_{q=\theta}^2 |\alpha_q^0\rangle) \otimes (\otimes_{q=1}^1 |\alpha_q^1\rangle) \otimes (\otimes_{p=\theta}^0 |c_p^0\rangle) \otimes (|w\rangle)\right). \tag{1-2}
 \end{aligned}$$

It is assumed that $(4 \times L \times L)$ quantum registers of θ bits, $(\otimes_{k=L}^1 (\otimes_{j=L}^1 (\otimes_{p=3}^0 (\otimes_{q=\theta}^1 |d2_{k,j,p,q}\rangle))))$, are used to encode the values of the X coordinate for four neighbors of the k th lattice site in a two-dimensional lattice for $1 \leq k \leq L$. Also it is supposed that $(4 \times L \times L)$ quantum registers of θ bits, $(\otimes_{k=L}^1 (\otimes_{j=L}^1 (\otimes_{p=3}^0 (\otimes_{q=\theta}^1 |b2_{k,j,p,q}\rangle))))$, are used to encode the values of the Y

coordinate for four neighbors of the k th lattice site in a two-dimensional lattice for $1 \leq k \leq L$. Their initial states for $(\otimes_{k=L}^1(\otimes_{j=L}^1(\otimes_{p=3}^0(\otimes_{q=\theta}^1|d2_{k,j,p,q}\rangle))))$ and $(\otimes_{k=L}^1(\otimes_{j=L}^1(\otimes_{p=3}^0(\otimes_{q=\theta}^1|b2_{k,j,p,q}\rangle))))$ are set to $(\otimes_{k=L}^1(\otimes_{j=L}^1(\otimes_{p=3}^0(\otimes_{q=\theta}^1|d2_{k,j,p,q}^0\rangle))))$ and $(\otimes_{k=L}^1(\otimes_{j=L}^1(\otimes_{p=3}^0(\otimes_{q=\theta}^1|b2_{k,j,p,q}^0\rangle))))$.

It is assumed that $(4 \times L \times L)$ quantum registers of $(2 \times \theta + 1)$ bits, $(\otimes_{k=L}^1(\otimes_{j=L}^1(\otimes_{p=3}^0(\otimes_{t=2 \times \theta}^0|R_{k,j,p,t}\rangle))))$, are used to store the result of comparing the value of the two-dimensional coordinate for the k th lattice site with the value of the two-dimensional coordinate for the j th lattice in a two-dimensional lattice for $1 \leq k \leq L$ and $1 \leq j \leq L$. The initial states for $(\otimes_{k=L}^1(\otimes_{j=L}^1(\otimes_{p=3}^0|R_{k,j,p,0}\rangle))))$ are set to $(\otimes_{k=L}^1(\otimes_{j=L}^1(\otimes_{p=3}^0|R_{k,j,p,0}^1\rangle))))$, and the initial states for $(\otimes_{k=L}^1(\otimes_{j=L}^1(\otimes_{p=3}^0(\otimes_{t=2 \times \theta}^1|R_{k,j,p,t}\rangle))))$ are set to $(\otimes_{k=L}^1(\otimes_{j=L}^1(\otimes_{p=3}^0(\otimes_{t=2 \times \theta}^1|R_{k,j,p,t}^0\rangle))))$. Also it is supposed that $(4 \times L \times L)$ quantum registers of two bits, $(\otimes_{k=L}^1(\otimes_{j=L}^1(\otimes_{p=3}^0(\otimes_{u=2}^1|\delta_{k,j,p,u}\rangle))))$, are applied to store the intermediate result for checking whether the k th lattice site and the j th lattice site are hydrophobic and they have loose contact in the corresponding conformation. The initial states for $(\otimes_{k=L}^1(\otimes_{j=L}^1(\otimes_{p=3}^0(\otimes_{u=2}^1|\delta_{k,j,p,u}\rangle))))$ are set to $(\otimes_{k=L}^1(\otimes_{j=L}^1(\otimes_{p=3}^0(\otimes_{u=2}^1|\delta_{k,j,p,u}^0\rangle))))$. The quantum registers above are auxiliary quantum registers and are used to figure out the global energy of each possible conformation from conformational space of a protein with an amino acid sequence $\mathbf{a} = a_1 \dots a_L$ in the hydrophobic–hydrophilic model in a two-dimensional lattice. **Algorithm 1-2** is applied to complete the function of figuring out the global energy for each possible conformation among conformation space. The notations used in **Algorithm 1-2** are defined in the previous subsections.

Algorithm 1-2: Figure out the global energy of each possible conformation from conformational space of a protein with an amino acid sequence $\mathbf{a} = a_1 \dots a_L$ in the hydrophobic–hydrophilic model in a two-dimensional lattice.

- (1) For $k = 1$ to L
- (2) For $j = (k + 2)$ to L
- (3) For $p = 0$ to 3

(3a) Based on that $(\otimes_{q=\theta}^1|d2_{k,j,p,q}\rangle)$ and $(\otimes_{q=\theta}^1|b2_{k,j,p,q}\rangle)$ are target bits and $(\otimes_{q=\theta}^1|d_{k,q}\rangle)$ and $(\otimes_{q=\theta}^1|b_{k,q}\rangle)$ are control bits, a unitary gate $(\otimes_{q=\theta}^1|d2_{k,j,p,q}^0 \oplus d_{k,q}\rangle) \otimes (\otimes_{q=\theta}^1|b2_{k,j,p,q}^0 \oplus b_{k,q}\rangle)$ is used to copy the

values of $(\otimes_{q=\theta}^1 |d_{k,q}\rangle)$ and $(\otimes_{q=\theta}^1 |b_{k,q}\rangle)$ from the quantum state vector $(\left(|\Phi_{k,j,p}^0\rangle \right) \otimes (\otimes_{u=2}^1 |\delta_{k,j,p,u}^0\rangle) \otimes ({}_{t=2\times\theta}^1 |R_{k,j,p,t}^0\rangle) \otimes (|R_{k,j,p,0}^1\rangle) \otimes (\otimes_{q=\theta}^1 |d2_{k,j,p,q}^0\rangle) \otimes (\otimes_{q=\theta}^1 |b2_{k,j,p,q}^0\rangle) \otimes (\left| \Omega_{6-(L+((k-1)\times(L-k-1)\times 4+(j-k-2)\times 4+p))} \right\rangle)$), and the following new quantum state vector is obtained:

$$\begin{aligned} & \left| \Omega_{6-(L+((k-1)\times(L-k-1)\times 4+(j-k-2)\times 4+p+1),1} \right\rangle = \left(\frac{1}{\sqrt{2^{2\times(L-1)}}} \right) \left(|\Phi_{k,j,p}^0\rangle \right) \otimes \\ & (\otimes_{u=2}^1 |\delta_{k,j,p,u}^0\rangle) \otimes ({}_{t=2\times\theta}^1 |R_{k,j,p,t}^0\rangle) \otimes (|R_{k,j,p,0}^1\rangle) \otimes \\ & \left(\sum_{w=0}^{2^{2\times(L-1)}-1} (\otimes_{q=\theta}^1 |d2_{k,j,p,q}^0 \oplus d_{k,q}\rangle) \otimes (\otimes_{q=\theta}^1 |b2_{k,j,p,q}^0 \oplus b_{k,q}\rangle) \right) \otimes \\ & (\otimes_{e=L}^1 (\otimes_{q=\theta}^1 |d_{e,q}\rangle) \otimes (\otimes_{q=\theta}^1 |b_{e,q}\rangle)) \otimes (\otimes_{q=\theta}^2 |\alpha_q^0\rangle) \otimes (\otimes_{q=1}^1 |\alpha_q^1\rangle) \otimes \\ & (\otimes_{p=\theta}^0 |c_p^0\rangle) \otimes (|w\rangle). \end{aligned}$$

(3b) If the value of p is equal to zero, then a unitary gate (a quantum adder), $((\otimes_{q=\theta}^2 (|b2_{k,j,p,q}\rangle + |\alpha_q^0\rangle)) \otimes (|b2_{k,j,p,0}\rangle + |\alpha_q^1\rangle))$, is used to figure out the two-dimensional coordinate of the *up* neighbor for the k^{th} lattice site from the quantum state vector $(\left| \Omega_{6-(L+((k-1)\times(L-k-1)\times 4+(j-k-2)\times 4+p+1),1} \right\rangle)$. If the value of p is equal to one, then a unitary gate (a quantum adder), $((\otimes_{q=\theta}^2 (|d2_{k,j,p,q}\rangle + |\alpha_q^0\rangle)) \otimes (|d2_{k,j,p,0}\rangle + |\alpha_q^1\rangle))$, is applied to figure out the two-dimensional coordinate of the *right* neighbor for the k^{th} lattice site from the quantum state vector $(\left| \Omega_{6-(L+((k-1)\times(L-k-1)\times 4+(j-k-2)\times 4+p+1),1} \right\rangle)$. If the value of p is equal to two, then a unitary gate (a quantum subtraction), $((\otimes_{q=\theta}^2 (|b2_{k,j,p,q}\rangle - |\alpha_q^0\rangle)) \otimes (|b2_{k,j,p,0}\rangle - |\alpha_q^1\rangle))$ is used to figure out the two-dimensional coordinate of the *down* neighbor for the k^{th} lattice site from the quantum state vector $(\left| \Omega_{6-(L+((k-1)\times(L-k-1)\times 4+(j-k-2)\times 4+p+1),1} \right\rangle)$. If the value of p is equal to three, then a unitary gate (a quantum subtraction), $((\otimes_{q=\theta}^2 (|d2_{k,j,p,q}\rangle - |\alpha_q^0\rangle)) \otimes (|d2_{k,j,p,0}\rangle - |\alpha_q^1\rangle))$, is used to figure out the two-dimensional coordinate of the *left* neighbor for the k^{th} lattice site from the quantum state vector

($\left| \Omega_{6-(L+((k-1)\times(L-k-1)\times 4+(j-k-2)\times 4+p+1),1)} \right\rangle$). Based on the condition of the execution for each iteration, the following new quantum state vector is obtained:

$$\begin{aligned} & \left| \Omega_{6-(L+((k-1)\times(L-k-1)\times 4+(j-k-2)\times 4+p+1),2)} \right\rangle = \left(\frac{1}{\sqrt{2^{2\times(L-1)}}} \right) \left(\left| \Phi_{k,j,p}^0 \right\rangle \otimes \right. \\ & \left. \left(\otimes_{u=2}^1 \left| \delta_{k,j,p,u}^0 \right\rangle \right) \otimes \left(\otimes_{t=2\times\theta}^1 \left| R_{k,j,p,t}^0 \right\rangle \right) \otimes \left(\left| R_{k,j,p,0}^1 \right\rangle \right) \otimes \right. \\ & \left. \left(\sum_{w=0}^{2^{2\times(L-1)}-1} \left(\otimes_{q=\theta}^1 \left| d2_{k,j,p,q} \right\rangle \right) \otimes \left(\otimes_{q=\theta}^1 \left| b2_{k,j,p,q} \right\rangle \right) \otimes \left(\otimes_{e=L}^1 \left(\otimes_{q=\theta}^1 \left| d_{e,q} \right\rangle \right) \otimes \right. \right. \\ & \left. \left. \left(\otimes_{q=\theta}^1 \left| b_{e,q} \right\rangle \right) \right) \otimes \left(\otimes_{q=\theta}^2 \left| \alpha_q^0 \right\rangle \right) \otimes \left(\otimes_{q=1}^1 \left| \alpha_q^1 \right\rangle \right) \otimes \left(\otimes_{p=\theta}^0 \left| c_p^0 \right\rangle \right) \otimes \left(|w\rangle \right) \right). \end{aligned}$$

(3c) Based on that $\left(\otimes_{q=\theta}^1 \left| d2_{k,j,p,q} \right\rangle \right)$ and $\left(\otimes_{q=\theta}^1 \left| b2_{k,j,p,q} \right\rangle \right)$ are target bits and $\left(\otimes_{q=\theta}^1 \left| d_{j,q} \right\rangle \right)$ and $\left(\otimes_{q=\theta}^1 \left| b_{j,q} \right\rangle \right)$ are control bits, a unitary gate $\left(\otimes_{q=\theta}^1 \left| d2_{k,j,p,q} \oplus d_{j,q} \right\rangle \right) \otimes \left(\otimes_{q=\theta}^1 \left| b2_{k,j,p,q} \oplus b_{j,q} \right\rangle \right)$ is used to judge whether the j th lattice site is one of four neighbors for the k th lattice site from the quantum state vector $\left(\left| \Omega_{6-(L+((k-1)\times(L-k-1)\times 4+(j-k-2)\times 4+p+1),2)} \right\rangle \right)$, and the following new quantum state vector is obtained:

$$\begin{aligned} & \left| \Omega_{6-(L+((k-1)\times(L-k-1)\times 4+(j-k-2)\times 4+p+1),3)} \right\rangle = \left(\frac{1}{\sqrt{2^{2\times(L-1)}}} \right) \left(\left| \Phi_{k,j,p}^0 \right\rangle \otimes \right. \\ & \left. \left(\otimes_{u=2}^1 \left| \delta_{k,j,p,u}^0 \right\rangle \right) \otimes \left(\otimes_{t=2\times\theta}^1 \left| R_{k,j,p,t}^0 \right\rangle \right) \otimes \left(\left| R_{k,j,p,0}^1 \right\rangle \right) \otimes \right. \\ & \left. \left(\sum_{w=0}^{2^{2\times(L-1)}-1} \left(\otimes_{q=\theta}^1 \left| d2_{k,j,p,q} \oplus d_{j,q} \right\rangle \right) \otimes \left(\otimes_{q=\theta}^1 \left| b2_{k,j,p,q} \oplus b_{j,q} \right\rangle \right) \otimes \right. \\ & \left. \left(\otimes_{e=L}^1 \left(\otimes_{q=\theta}^1 \left| d_{e,q} \right\rangle \right) \otimes \left(\otimes_{q=\theta}^1 \left| b_{e,q} \right\rangle \right) \right) \otimes \left(\otimes_{q=\theta}^2 \left| \alpha_q^0 \right\rangle \right) \otimes \left(\otimes_{q=1}^1 \left| \alpha_q^1 \right\rangle \right) \otimes \right. \\ & \left. \left(\otimes_{p=\theta}^0 \left| c_p^0 \right\rangle \right) \otimes \left(|w\rangle \right) \right). \end{aligned}$$

(3d) A unitary gate, $\left(\left| \Phi_{k,j,p}^0 \oplus \delta_{k,j,p,2} \right\rangle \right) \otimes \left(\left| \delta_{k,j,p,2}^0 \oplus (s_j \bullet \delta_{k,j,p,1}) \right\rangle \right) \otimes$
 $\left(\left| \delta_{k,j,p,1}^0 \oplus (s_k \bullet R_{k,j,p,2\times\theta}) \right\rangle \right) \otimes$
 $\left(\otimes_{t=2\times\theta}^{\theta+1} \left| R_{k,j,p,t}^0 \oplus \overline{(d2_{k,j,p,t-\theta} \bullet R_{k,j,p,t-1})} \right\rangle \right) \otimes$
 $\left(\otimes_{t=\theta}^2 \left| R_{k,j,p,t}^0 \oplus \overline{(b2_{k,j,p,t} \bullet R_{k,j,p,t-1})} \right\rangle \right) \otimes$
 $\left(\left| R_{k,j,p,1}^0 \oplus \overline{(b2_{k,j,p,1} \bullet R_{k,j,p,0}^1)} \right\rangle \right)$, is applied to check whether the k th lattice site and the j th lattice site are hydrophobic and they have loose contact in the

corresponding conformation from $(|\Omega_{6-(L+((k-1)\times(L-k-1)\times 4+(j-k-2)\times 4+p+1),3)}\rangle)$, and the following new quantum state vector is obtained:

$$\begin{aligned} & \left| \Omega_{6-(L+((k-1)\times(L-k-1)\times 4+(j-k-2)\times 4+p+1)} \right\rangle = \\ & \left(\frac{1}{\sqrt{2^{2\times(L-1)}}} \right) \left(\sum_{w=0}^{2^{2\times(L-1)}-1} (|\Phi_{k,j,p}\rangle) \otimes (\otimes_{u=2}^1 |\delta_{k,j,p,u}\rangle) \otimes (\otimes_{t=2\times\theta}^1 |R_{k,j,p,t}\rangle) \otimes \right. \\ & \left. (|R_{k,j,p,0}\rangle) \otimes (\otimes_{q=\theta}^1 |d2_{k,j,p,q}\rangle) \otimes (\otimes_{q=\theta}^1 |b2_{k,j,p,q}\rangle) \otimes (\otimes_{e=L}^1 (\otimes_{q=\theta}^1 |d_{e,q}\rangle)) \right. \\ & \left. \otimes (\otimes_{q=\theta}^1 |b_{e,q}\rangle) \otimes (\otimes_{q=\theta}^2 |\alpha_q^0\rangle) \otimes (\otimes_{q=1}^1 |\alpha_q^1\rangle) \otimes (\otimes_{p=\theta}^0 |c_p^0\rangle) \otimes (|w\rangle) \right). \end{aligned}$$

(3e) EndFor

(2a) EndFor

(1a) EndFor

EndAlgorithm

Lemma 1-2: Algorithm 1-2 is used to compute the global minimum energy of each possible conformation from conformational space of a protein with an amino acid sequence $\mathbf{a} = a_1 \dots a_L$ in the hydrophobic–hydrophilic model in a two-dimensional lattice.

1.3 Introducing Quantum Networks for Finding Global Minimum Energy from Conformational Space to a Protein in the Hydrophobic-Hydrophilic Model in the Two-Dimensional Lattice

It is assumed that for $1 \leq k \leq L$, $0 \leq p \leq 3$, and $0 \leq j \leq ((k-1) \times 4 + p)$, $|z_{k,p,j+1}\rangle$ is applied to store the result of those conformations in conformational space that have $(j+1)$ ones, and $|z_{k,p,j}\rangle$ is used to store the result of these conformations in conformational space that have j ones after the influence of $(|\Phi_{k,1,p}\rangle)$ through $(|\Phi_{k,L,p}\rangle)$ to the number of ones is figured out by U_{CONE} . Lemma 1-3 is used to describe that the parallel logic computation is completed by U_{CONE} .

Lemma 1-3: The parallel logic computation performed by a unitary gate, U_{CONE} , is (1.2):

(1) For $k = 1$ to L

(2) For $p = 0$ to 3

(3) For $j = ((k-1) \times 4 + p)$ down to 0

(3a) If $(p = 0)$ Then

(3b) $|z_{k,p,j+1}\rangle = |z_{k,p,j+1}\rangle \oplus ((\sqrt{L-1} |\Phi_{k,i+1,p}\rangle) \wedge (|z_{k-1,3,j}\rangle \wedge$

$(\wedge_{x=j+2}^{4\times(k-1)+p+1} |z_{k,p,x}\rangle)))$, and $|z_{k,p,j}\rangle = |z_{k,p,j}\rangle \oplus ((\sqrt{L-1} |\Phi_{k,i+1,p}\rangle) \wedge |z_{k-1,3,j}\rangle)$.

(3c) **Else**

$$(3d) \left| z_{k,p,j+1} \right\rangle = \left| z_{k,p,j+1} \right\rangle \oplus \left(\left(\bigvee_{i=0}^{L-1} \left| \Phi_{k,i+1,p} \right\rangle \right) \wedge \left(\left| z_{k,p-1,j} \right\rangle \wedge \left(\bigwedge_{x=j+2}^{4 \times (k-1) + p+1} \overline{\left| z_{k,p,x} \right\rangle} \right) \right) \right), \text{ and } \left| z_{k,p,j} \right\rangle = \left| z_{k,p,j} \right\rangle \oplus \left(\left(\bigvee_{i=0}^{L-1} \overline{\left| \Phi_{k,i+1,p} \right\rangle} \right) \wedge \left| z_{k,p-1,j} \right\rangle \right).$$

(3e) **EndIf**

(3f) **EndFor**

(2a) **EndFor**

(1a) **EndFor**

Lemma 1-3(a): A unitary gate, U_{CONE} , in **Lemma 1-3** can be implemented by means of $(32 \times L^3 + 36 \times L^2 - 8 \times L)$ **NOT** gates, $(16 \times L^2 + 4 \times L)$ **CNOT** gates, and $(8 \times L^3 + 62 \times L^2)$ **CCNOT** gates.

Parallelization of Wu's Method with Multithreading*

Hongbo Li and Suping Wu**

School of Mathematics and Computer Science
Ningxia University
Yinchuan, China
lihongbo520066@163.com, wusup@263.net

Abstract. The establishment of Wu's method founded an integrated theory and offered an efficient algorithm for solving polynomial equation systems. This paper focuses on the characteristic series algorithm, the core of Wu's method. Wu's method, based on symbolic computation, is very compute-intensive. This point usually leads to a time-consuming and inefficient computing process. So parallel computing is introduced to accelerate the computing process. In this paper the corresponding parallel algorithm is presented and its parallel implementation based on multithreading in Maple system is given. The experiment demonstrates a considerable speedup, which shows the high efficiency of this parallel algorithm.

Keywords: Wu's Method, Parallelization, Characteristic Series, Multithreading.

1 Introduction

The problem of solving nonlinear algebraic equation systems has always been one of the most fundamental computing problems. Wu's method offers an efficient algorithm for this problem, which is a substantial progress in the research of modern nonlinear mathematics. However, Wu's method based on symbolic computation, which is quite different from numerical computing, is very compute-intensive [1,2]. This is the very point leading to a time-consuming and inefficient computing process. High speed is therefore essential to use this method in practical application. So parallel computing is introduced to improve the performance of Wu's method.

Some achievement has been made in the research of parallelization for computing characteristic series, denoted by CSS, which is the core of Wu's method. Wang D. M. [3,4] researched on the parallelism of CSS in Maple system by a local network; Ajwa I. et al. [5,6] parallelized this problem in the environment of PVM and SACLIB. Wu Y. W. et al. [7] implemented the parallelization of CSS based on MPI and ELIMINO; Wu S. P. [8] has made more beneficial explorations based on distributed Maple system.

All these researches on the parallelization of Wu's method are based on distributed system. Multicore processors are now commonplace and the serial programs cannot

* Supported by the National Natural Science Foundation of China under Grant No. 60963004.

** Corresponding author.

make full use of the computing resources. To counter this problem, this paper gives a parallel algorithm and implementation of CSS with multithreading in Maple on the platform of multicore computer. The experiment displays that a preferable speedup is gained by the parallel algorithm.

2 Wu’s Method

To solve polynomial equation systems with Wu’s method is very similar to solve linear ones with Gauss elimination method. Their essential steps are to simplify the original polynomials to a triangular form. The objective of Wu’s method is to use appropriate triangular form of polynomials to precisely describe the zero set of a given polynomial equation system. Characteristic set is just the appropriate triangular form, so the problem of computing the zero set turns into the issue of obtaining the relevant characteristic series CSS, i.e. the union of all characteristic sets which can depict all the zero points.

2.1 The Main Theorems [1]

— Theorem 1 (Well Ordering Principle)

Given a finite, non-empty set PS of non-zero polynomials, $PS=\{P_1, P_2, \dots, P_m\}$, $P_i \in K[X]$, $i=1, 2, \dots, m$, there certainly exists a mechanized algorithm, with which after limited steps of computation a basic set of PS , denoted by BS , can be computed, which satisfies that the pseudo remainder of each polynomial in PS with respect to BS is zero. If this BS is non-contradictory, then it is one characteristic set of PS , denoted by CS .

— Theorem 2 (Zero Set Structure Theorem)

Given the characteristic set CS , $CS=\{C_1, C_2, \dots, C_r\}$, of polynomial set PS , $PS=\{P_1, P_2, \dots, P_m\}$. The initial of C_i is I_i , $i=1, 2, \dots, r$. The zero set, of PS , denoted by $Zero(PS)$ has this structure:

$$Zero(PS) = Zero(CS / I) \bigcup_{i=1}^r Zero(PS, I_i), I = \prod_{i=1}^r I_i . \tag{1}$$

— Theorem 3 (Zero Set Decomposition Theorem)

Given the polynomial set PS , $PS=\{P_1, P_2, \dots, P_m\}$, whose zero set can be decomposed into a union set of a series of characteristic sets’ zero sets:

$$Zero(PS) = \bigcup_j Zero(CS_j / I_j) , \tag{2}$$

The set of index j in this equation is finite, I_j represents the product of all the initials in CS_j .

The well ordering principle is applied in the process of computing every specific CS ; the zero set structure theorem and the zero set decomposition theorem describe the constructive description of zero set.

2.2 Sequential Algorithm of Characteristic Series

The sequential algorithm in this paper is extracted in the Wsolve package [9] in Maple created by Wang D. K. It consists of several important parts: computing basic sets, pseudo division of polynomials, factorization, calculating characteristic sets, decomposition for computing characteristic series. For the sake of simplicity, the last part, basis for parallel algorithm description followed, is given only. Its sequential algorithm is put as follows.

```

Wsolve ( Input: PS;  Output: CSS )
Begin
  // Decomposing PS into many independent polynomial
  //sets, QS is a set of sets
  QS := Factorization (PS)
  while PSET in QS do
    CSS:= CSS∪CharacteristicSetSeries(PSET)
  end while
End

// Computing the characteristic series of PS
CharacteristicSetSeries( Input: PS;  Output: CSS )
Begin
  QS := PS
  CS := CharacteristicSet(PS)
  // The difference set between CS and constant set
  IS := init(CS) -{constant}
  while I in IS do
    // CharacteristicSet: computing CS
    CSS := CS∪CharacteristicSet(QS∪{I})
  end while
End

```

3 Parallel Algorithm

3.1 Task Programming Model in Maple

Task Programming Model [10] is firstly introduced in Maple 13, which is a high level programming tool that simplifies many aspects of parallel programming. It enables parallelism by executing multiple tasks within a single process. In this model, tasks share memory thus they can work closely together, sharing data with low overhead.

The following is a brief description of how the Task Programming Model works. Internally, Maple creates a pool of threads, where the number of threads is equal to the value of *kernelopts(numcpus)*. By default this is the number of processors available on the computer. This pool is created during the first call to `Threads:-Task:-Start`. Each thread can execute one task at a time. As tasks are created they are stored in Task Queue. When a thread completes a task, it takes another task from Task Queue. Task Queue, storing tasks, contains a double ended queue for each thread. By default a thread will always execute the task that is at the head of its queue. Further, new tasks are added to the current thread's queue. If a thread finds that its queue is empty, it will try to steal a task from another thread's queue. When stealing a task, the task is removed from the tail of the queue.

3.2 Parallel Algorithm Design

Factorization not only is significantly vital to simplify polynomial set, but also plays an important role in symbolic computation. In the process of solving polynomial equation systems, it is a common phenomenon that there are usually polynomials can be factorized, especially meeting large-scale data input. So under this situation factorization facilitate the computation greatly. As the algorithm given above, to compute CSS is to factorize the polynomial set firstly, in which way the zero set of original polynomial equation system can be converted into the union of a series of polynomial equations' zero sets.

Take an example, given polynomial set (A, B) , if A and B can be factorized in this way: $A=A_1 \times A_2$, $B=B_1 \times B_2$, then the original zero set could be decomposed into 4 zero sets (2×2) as follows:

$$\text{Zero}(A, B) = \text{Zero}(A_1, B_1) \cup \text{Zero}(A_1, B_2) \cup \text{Zero}(A_2, B_1) \cup \text{Zero}(A_2, B_2). \quad (3)$$

However, after factorization a problem below appears. Given polynomial set $PS = \{P_1, P_2, \dots, P_m\}$, suppose that P_1, P_2, \dots, P_m can be respectively factorized in to i_1, i_2, \dots, i_m factors, $i_k \geq 1$, $k = 1, 2, \dots, m$. Accordingly the zero set problem of PS can be decomposed into completely the same problems. The number of these problems is $i_1 \times i_2 \times \dots \times i_m$. Namely a single zero set computing problem is transformed into a large number of the same problems. If these problems is solved by serial program one by one, it will surely consumes much time. On the contrary, these independent problems can be computed by parallel program, which will surely shorten the computing time.

So this part can be parallelized to improve the computation efficiency greatly. Parallel algorithm is described as follows.

```

Wsolve (Input: PS; Output: CSS)
Begin
  QS:= Factorization (PS)
  while PSET in QS par-do
    CSS:= CSS ∪ CharacteristicSetSeries( PSET )
  odrap
End

```

3.3 Parallel Implementation

When programming with multithreading, in order to get a good scalability it is so vital for us to consider how to divide the workload seriously to achieve a state of load balancing and the full use of computing capabilities. The common approach is to divide the input data into several even parts, whose number is equal to the computing cores'. However, this surely doesn't mean the workload is split evenly, because the evenly split input data may contains different computational complexity, i.e. workload. One approach to solve this problem is to create a large number of small tasks. In this way even if one task requires more time to run, the other cores can run many other tasks while one core is running the long task. One limitation is that creating tasks requires much more resources and overhead.

In parallel programming the parallel granularity has to be suitable to achieve a good state of load balancing. Here two tactics are used: granularity control and recursive interlaced bisection.

(1) Granularity control mainly contributes to divide the input data evenly. Hereon we design a granularity control formula:

$$granularity := \text{ceil} \left(\frac{scale(inputdata)}{k \cdot numcpu} \right). \tag{4}$$

Function note: *granularity*, upper limit of granularity for each task; *inputdata*, input; *numcpu*, the number of cores in CPU; *k*, control parameter; *scale()*, function to compute the input scale; *ceil()*, top integral function.

After analyzing the input scale and the number of computing cores, the function can regulate through control parameter *k* to achieve a good scalability. With this function, the input scale is divided into $k \times numcpu$ parts evenly.

(2) In order to ensure the workload in each part divided by the upper formula is as even as possible, the recursive interlaced bisection method is imported. This tactic could ease the load imbalancing problem resulting from the different workload between evenly divided parts. Recursive bisection based on divide and conquer [11] can be put like this: split the given domain on one-dimensional direction; then split in the two subdomains recursively till each subdomain meets the demand of granularity. The interlaced bisection can be described in this way: suppose there is a given task, $task=[task_1, task_2, \dots, task_{2n}]$, after disposition with interlaced bisection method the two child-tasks should be like this, $childtask_1=[task_1, task_3, \dots, task_{2n-1}]$, $childtask_2=[task_2, task_4, \dots, task_{2n}]$. The combining method of these two ways is the so called recursive interlaced bisection strategy, which we need.

The following is the procedure of parallel implementation with multithreading.

(1) Factorize the given polynomial set into a group of polynomial sets. The result is a set of polynomial sets, denoted by *pset*.

(2) Setup control parameter *k* to get the *granularity* you want by computing the granularity control function. The parameter *numcpu* can be obtained from the function in Maple, which is *kernelopts(numcpus)*. These functions used in the granularity control function.

(3) Start task with *pset* using *Threads[Task][Start]*.

(4) Decide whether or not to split the current task, by comparing the scale of current task and the *granularity* you set. If segmentation is needed, then recursive interlaced bisection method is used to split the task. So you get 2 data groups. Two child tasks with the divided data groups can be started with *Threads[Task][Continue]*. On the contrary, if segmentation is not needed, we compute the characteristic series of polynomial sets in this task.

(5) If child tasks are started in the previous step, repeat step (4) till each task, whose scale is smaller than granularity, are executed and returns with a result.

One thing to note is that the results returned from tasks are disposed and assembled by a function. And this function is a parameter in *Threads[Task][Continue]*.

4 Experimental Results and Analysis

4.1 Experimental Results

Experiment has been performed on two different platforms with sequential program and multithreading parallel program so as to testify the efficiency of parallel algorithm. Sequential program is the Wsolve package created by Wang D. K. and parallel program is modified based on this package. The two different platforms have the same software environment but different hardwares. The specific environment is given below.

Software environment: Windows XP, Maple 14.

Hardware platform: (1) dual-core platform, Intel(R) Core(TM) 2 Duo CPU T6570; (2) quad-core platform, Intel(R) Core(TM) i7 950.

Besides, the given problems used in the experiment are obtained by enlarging several small but basic polynomial sets. The enlargement method is to increase the number and power of variables, and the number of polynomials.

Table 1 shows the experimental results which gives the speedup ratio for different problems under two platforms. Execution time is given in seconds.

Table 1. Experimental Results

Example		1	2	3	4	5	6
Dual-core	<i>Serial time</i>	0.219	5.504	8.736	16.556	45.396	53.414
	<i>Parallel time</i>	0.108	3.524	5.975	9.432	30.904	35.958
	<i>Speedup</i>	2.028	1.562	1.462	1.747	1.469	1.485
Quad-core	<i>Serial time</i>	0.079	2.938	4.954	9.547	27.500	38.250
	<i>Parallel time</i>	0.062	2.672	1.798	3.438	11.500	13.421
	<i>Speedup</i>	1.274	1.100	2.755	2.778	2.391	2.855

After the workload reaches a certain scale and the speedup gets a certain stable state, the average speedup is gained by computing a mean value of several stable speedups. The average speedup under dual-core platform is the mean value of the last 5 speedups; under quad-core platform, the last 4. The average speedup ratio under the first platform is 1.531; the second one, 2.695.

4.2 Analysis

Based on the above experimental results, some phenomena can be concluded. (1) Under dual-core platform, the first problem's speedup is higher than the others'; (2) Under quad-core platform, the first two problems' speedup is lower than the others' obviously; (3) As the scale of problems growing larger, the speedup under both platforms becomes stable in fluctuation gradually; (4) After the problems' scale get a certain degree, the average speedup under dual-core platform is about 1.531; under quad-core platform, 2.695.

Targeting these phenomena above, in-depth analysis is done below respectively. (1) The first problem's scale is the smallest. Phenomenon 1 roots in that the first one's load balancing is better than others'; (2) The reason for phenomenon 2 is that the number of cores increase while the problem scale stays, which leads to that the ratio of overhead is too large. So if advancing speedup, the problem scale must be enlarged. (3) The reasonable cause of phenomenon 3 can be interpreted below. According to speedup formula the speedup converges to a definite value when the problem scale grows while the number of CPU cores and overhead (overhead is a function on the number of cores, while irrelevant with the problem scale) stay. Due to that absolute load balancing cannot be reached, fluctuation cannot be avoided in the converging process.

5 Conclusion

The new parallel algorithm given in this paper is grounded on the condition that the given polynomial set can be factorized. Large-scale polynomial sets often contains many factorizable polynomials, which facilitates the algorithm greatly. Because a single zero set computing problem is transformed into a large number of the same problems after factorization, the parallel algorithm is very suitable under the condition of meeting large-scale input and frequent factorizations. Surely limitations cannot be avoided too. That is when the problem scale is not so big or factorizable polynomials are scarce, this algorithm fail to meet the high performance. So aiming at cracking this trouble, further research is needed. There are some points worthy of our attention, such as the ratio of parallelizable part in workload, overhead for parallelization, finding more parallelizable factors and combining them etc. All these considerations rely on further research and exploration.

Acknowledgment. This paper is funded by the National Natural Science Foundation of China (No. 60963004). We would like to express our sincere gratitude to the National Natural Science Foundation for the financial support. And many thanks to all those people who supported us.

References

1. Shi, H.: Introduction to Mechanized Mathematics. Hunan Education Press, Changsha (1988) (in Chinese)
2. Wu, W.T.: Mathematics mechanization: its origin, status and prospect. *J Sys. Sci. & Math. Scis* 28, 898–904 (2008) (in Chinese)
3. Wang, D.M.: Characteristic sets and zero structure of polynomial sets, Lecture Notes, RISC-Linz, Johannes Kepler University, Austria (1989)
4. Wang, D.M.: On the parallelization of characteristic-set-based algorithms. In: Proceedings of the 1st International Conference ACPC, Salzburg, Austria, pp. 338–349 (1991)
5. Ajwa, I.: Parallel algorithms and implementations for the Gröbner bases algorithm and the characteristic sets method, Ph.D. Dissertation, Kent State University, Kent, Ohio (December 1998)
6. Ajwa, I., Wang, P., Lin, D.: Another attempt for parallel computation of characteristic sets. In: Proceedings of the Fourth Asian Symposium on Computer Mathematics (ASCM 2000). Lecture Notes, Series on Computing, vol. 8, pp. 63–66 (2000)
7. Wu, Y.W., Yang, G.W., Yang, H., Zheng, W.M., Lin, D.D.: A distributed computing model for Wu's method. *Journal of Software* 16, 384–391 (2005)
8. Wu, S.P.: Research On parallelization of the related characteristic set method based on distributed Maple system. *Modern Computer*, 14–16 (October 2009) (in Chinese)
9. Wang, D.K.: A Maple package for solving system of polynomial equations, *Mathematics-Mechanization Re-search Preprints*, pp. 11–14 (October 1993)
10. Maplesoft Online Help, Task Programming Model, <http://www.maplesoft.com/support/help>
11. Chen, G.L.: Parallel computing — structure, algorithm and programming, 3rd edn. Higher Education Press, Beijing (2003) (in Chinese)

The Influence of the Sectional form of Labyrinth Emitter on the Hydraulic Properties

Zhiqin Li^{1,2,*} and Lin Li³

¹Xi'an University of Technology, Xi'an, Shaanxi 710048, China

²Taiyuan University of Technology, Taiyuan, Shanxi 030024, China

lzc_lzc@163.com

³Taiyuan Survey and Design Institute of Water Conservancy, Taiyuan, Shanxi 030002, China

Abstract. This paper studied the influences of the sectional form of the labyrinth emitter used for single wing labyrinth drip irrigation hose on the hydraulic properties; analyzed theoretically the relationship between flow index of the emitter and flow pattern and the influences of the sectional form of the emitter on the flow index; simulated the flow field of the emitters with 5 different sectional forms under the conditions of 14 different discharges. The following conclusions are obtained: 1) To the non-compensatory drip emitter, the minimum flow index is 0.5 for completely turbulent flow; 2) Both discharge coefficient and flow index of the labyrinth emitter are positively correlation with the wetted perimeter. The bigger the wetted perimeter is, the higher the sensitivity of the emitter discharge to the pressure is. 3) With equivalent labyrinth path cross section area an emitter with a square shaped cross section performs better than one with a rectangle section.

Keywords: Hydraulics, drip irrigation, labyrinth emitter.

1 Introduction

China is a large agricultural country, 73% of the total water resources are used for agriculture. Being lacking of water resources, water shortage has become the largest restricting factors for agricultural development. Therefore, it is very necessary and urgent to develop the water-saving irrigation and popularize advanced water saving irrigation technology. Drip irrigation is one of the most effective methods among all water-saving irrigation technologies [1-2], and emitter is the key part of this technology. The efficiency of drip irrigation is mainly determined by emitter's hydraulic performance.

The single wing labyrinth drip irrigation hose, one of the drip irrigation equipments, gets wide application especially in Xinjiang Province due to its lower manufacturing costs. Combined with plastic film mulching cultivation techniques, it basically meets the request of farm crops drip irrigation by its low cost for its installation and renewal.

* Corresponding author.

As a result, it pioneered a new way to develop the efficient and water-saving irrigation technologies for drought areas. The emitter of this drip irrigation hose adopts labyrinth path. Its main processing method is to extrude the plastic pipe through the extruder, then squeeze it, and press it by special former before it cools, At the same time the final labyrinth is formed using vacuum clamshell mode[3-4]. On the point of production technology and products manufacturing precision, labyrinth flow passage with rectangular cross-section offers better precision control in processing.

At present the research about the factors influencing the hydraulic performance of labyrinth water device is mostly on labyrinth path structural parameters, such as the dental width, space between dental bottom and dental height[5-6]. In this paper, we focused on the emitter of the single wing labyrinth drip irrigation hose, and analyzed the influences of the section form of labyrinth emitter on the hydraulic performance.

2 Theoretical Analysis

2.1 Relationship between Flow Pattern Index and Flow Regime of the Emitter

According to Reynolds experiment results, when flow in pipelines is laminar, its energy loss h_w is directly proportional to v (cross-section average velocity); when flow in pipelines is turbulent, its energy loss h_w is proportional to $v^{1.75-2}$, when flow in pipelines is completely turbulent, energy loss h_w is proportional to v^2 . The general expression is:

$$h_w = \alpha v^m. \tag{1}$$

When the inlet area and the outlet area of emitter are equal, as its export pressure is atmospheric pressure, h_w is the pressure head of the inlet section. Then

$$h_w = \frac{P_1}{\gamma}. \tag{2}$$

Substituting (2) into (1), we have

$$\frac{P_1}{\gamma} = \alpha v^m. \tag{3}$$

Introducing sectional area A , $v=Q/A$, and replace the pressure head $\frac{P_1}{\gamma}$ with H , we obtain

$$H = \frac{\alpha}{A^m} Q^m. \tag{4}$$

Let $\frac{(\alpha)^{\frac{1}{m}}}{A} = \frac{1}{k}$, it gives

$$Q = kH^{\frac{1}{m}}. \tag{5}$$

Let $x=1/m$, it follows

$$Q = kH^x. \quad (6)$$

The formula (6) expresses the relationship between discharge and inlet pressure of the emitter, in which, x is the flow index of the emitter.

When flow is laminar, $m=1$, then, $x=1$; When flow is turbulence, $m=1.75\sim 2$, then, $x=0.57\sim 0.5$; When flow is of completely turbulent, $m=2$, then, $x=0.5$.

The smaller the flow index of the emitter is, the lower the pressure sensitivity is, which means the emitter has better hydraulic performance. From corollary above, for the non-compensatory emitter, when its internal flow is completely turbulent, the minimum flow pattern index is 0.5.

2.2 The Influence of Sectional Form of Emitter on Flow Index

The shorter the wetted perimeter of labyrinth path is, the smaller the resistance to the water of the emitter is, under the same conditions, the flow velocity is greater. When flow velocity increases up to a certain value, the flow pattern will become a completely turbulent flow, the flow pattern index of emitter gets to the minimum, then hydraulic performance of the emitter is improved, and the performances of anti-clogging and irrigation evenness are enhanced, too.

In view of the production technology and manufacture accuracy of labyrinth emitter of single wing labyrinth drip irrigation hose, the rectangular cross-section type is easier to manufacture than ones with other cross-section types.

The wetted perimeter of rectangular cross-section can be expressed as:

$$X = 2 (a + b). \quad (7)$$

Where, X is flow cross-section wetted perimeter, a and b are side lengths of the cross-section of the flow path respectively.

Derivation the formula above to side length of a , and let its derivative equal 0, and let:

$$\frac{dA}{da} = a \frac{db}{da} + b = 0. \quad (8)$$

gives $a=b$. It means that the side lengths of emitter are the same, i.e. when the sectional form of the emitter is square, the resistance of the emitter to the water is the least.

To sum up, taking square as emitter section shape will help to improve the hydraulic performance.

3 Numerical Simulation Analysis

Using Computational Fluid Dynamics(CFD) software – FLUENT to simulate flowing in emitters of 5 different cross-section shapes, and obtain the relationship equation between inlet pressure and discharge, so as to analyze the influences of sectional form of emitter on the flow coefficient and the flow pattern index.

3.1 Computational Models and Meshing

Five emitters with same length of flow path but different cross-section shape were simulated, and all 5 emitters are with sectional area of 1mm^2 , section sizes are listed in table 1.

GAMBIT, the preprocessor of FLUENT, is used to establish a 3d calculating model, meshing of wall part adopts boundary layer grid, meshing of internal uses hexahedral grid. Boundary layer was set with 6~8 layers, the first layer is with a height of 0.01mm. From the second layer, the height of each layer is 1.5 times of that of the former layer. Interior mesh dimensions are set by 0.1mm. Computational models and meshing are shown in fig.1.

Table 1. Computational models and section sizes

Serial number	Model number	Section width(mm)	Section height (mm)	Section wetted perimeter(mm^2)	Sectional area(mm)	Length of path(mm)
1	G1	1	1	4	1	301
2	G2	1.25	0.8	4.1	1	301
3	G3	1.6	0.625	4.45	1	301
4	G4	2	0.5	5	1	301
5	G5	2.5	0.4	5.8	1	301

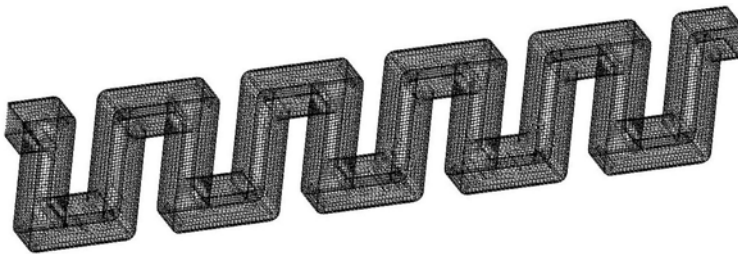


Fig. 1. Computational models and Meshes

3.2 Control Equations and Boundary Conditions

Control equations of numerical simulation include continuity equation and equations of motion.

Continuity equation:

$$\text{div}(\rho u) = 0. \tag{9}$$

Equations of motion:

$$\text{div}(\rho uu) = \text{div}(\mu \text{grad } u) - \frac{\partial p}{\partial x} \tag{10}$$

$$\text{div}(\rho vu) = \text{div}(\mu \text{grad } v) - \frac{\partial p}{\partial y} \tag{11}$$

$$\text{div}(\rho wu) = \text{div}(\mu \text{grad } w) - \frac{\partial p}{\partial z} \tag{12}$$

The calculation standard $k-\varepsilon$ turbulence model and SIMPLE algorithm are used to simulate the flow field. The constants in $k-\varepsilon$ turbulence model are:

$$C_1=1.44, C_2=1.92, C_\mu=0.09, \sigma_k=1.0, \sigma_\varepsilon=1.3.$$

Inlet boundary conditions are of velocity conditions, outlet boundary conditions are of outflow conditions, others are wall boundary conditions, and the wall is treated by standard wall function.

3.3 Calculation Results and Analysis

For the above five different labyrinth drip emitters, the internal flow fields of emitters under the conditions of 14 different discharges (1.436L/h, 1.616L/h, 1.796L/h, 1.976L/h, 2.156L/h, 2.33 L/h, 2.516L/h, 2.696L/h, 2.876L/h, 3.053L/h, 3.233L/h, 3.413L/h, 3.593L/h and 3.77L/h) are simulated. And the relationship curves between inlet pressure and discharge of emitters, and flow pattern index and discharge coefficient of emitters are obtained (see fig. 2 and table 2).

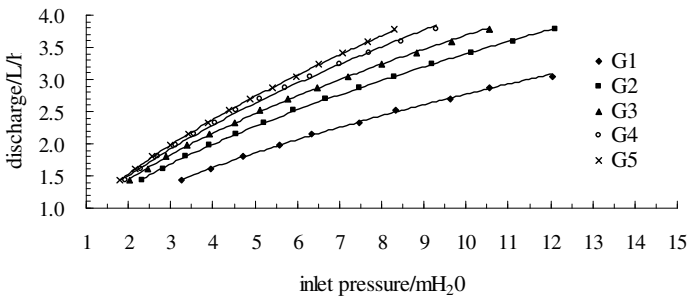


Fig. 2. The relationship curves between inlet pressure and discharge of emitters under 5 different sectional forms

From Table 2 we can know that 1) The discharge coefficient and flow path cross-section of the labyrinth emitters have a positive correlation with the wetted perimeter of the labyrinth path, the greater the wetted perimeter is, the greater the flow coefficient is, which indicates that the emitter flow is more sensitive to the variation of

Table 2. Discharge coefficient and flow pattern index of 5 emitters with different sections

Serial number	Emitter number	Wetted perimeter	Inlet pressure ~ discharge equation	Discharge coefficient	Flow pattern index
1	G1	4	$Q = 0.731H^{0.579}$	0.731	0.579
2	G2	4.1	$Q = 0.889H^{0.582}$	0.889	0.582
3	G3	4.45	$Q = 0.964H^{0.583}$	0.964	0.583
4	G4	5	$Q = 0.977H^{0.615}$	0.977	0.615
5	G5	5.8	$Q = 0.989H^{0.632}$	0.989	0.632

the pressure; 2) The flow index of labyrinth emitter has a positive correlation with the wetted perimeter of the labyrinth path, the greater the wetted perimeter is, the greater the flow index is, which shows that the sensitivity of the emitter flow to pressure increased; 3) The flow pattern index of the G1-type labyrinth emitter, which has smaller wetted perimeter, is less than those of the other 4 types, and is close to 0.57, indicating that flow regime in the emitter flow path is turbulent, consistent with the theoretical analysis.

4 Conclusions

Through analyzing and simulating the relationship between water flow pattern in the labyrinth path and flow coefficient and flow pattern index, and influences of the sectional form of the emitter on the flow coefficient and the flow pattern index, here are some conclusions.

- 1) Non-compensatory emitter has its minimum flow index of 0.5 when flow in the labyrinth passage is of full turbulence;
- 2) The discharge coefficient and flow pattern index of labyrinth emitter have a positive correlation with the wetted perimeter of the labyrinth path, the greater the wetted perimeter is, the higher the sensitivity of the emitter flow to the pressure is;
- 3) The theoretical analysis and numerical simulation results showed that when the section areas of the labyrinth path were the same, the hydraulic performance of emitter with square sectional shape is better than that of the one with rectangle section shape.

Acknowledgement. This work is supported by the Natural Science Foundation of Shanxi Province, China. (No. 2008011054).

References

1. Teng, L., Nie, J.: Development Status and Prospect of Water-saving Irrigation Technique in China. *Water Resources and Hydropower Engineering*. 28(3), 34–52 (1997)
2. Liu, Y.: Development status and prospect of drip irrigation for farm crops in china. *Water Saving Irrigation* 4, 26–31 (1998)

3. Zheng, Y.: The Selection of Dripper's Technical Performance Parameters and Structure in China. *Journal of Irrigation and Drainage* 11(2), 8–13 (1992)
4. Liu, J., Wei, Q., Lu, G., et al.: Overview of the development of production line technology on Drip irrigation belt. *Water Saving Irrigation* 7, 40–43 (2009)
5. Chang, Y., Niu, Q., Wang, W.: Numerical simulation and flow analysis of internal fluid in Drip irrigation emitters. *Journal of Northwest A&F University (Natural Science Edition)* 37(2), 203–208 (2009) (in Chinese)
6. Yu, L., Wu, P., Niu, Q., et al.: The Influence of the corner of Labyrinth Flow on Hydraulic performance of the Douche. *Transactions of the Chinese Society for Agricultural Machinery* 40(2), 63–67 (2009) (in Chinese)

Neural Network and Support Vector Machines in Slime Flotation Soft Sensor Modeling Simulation Research

Ranfeng Wang

Mineral processing engineering department, Taiyuan University of Technology TaiYuan, China
wrf197010@163.com

Abstract. The flotation process refined coal ash soft measuring is the key technology to the flotation process automation .Based on the generalized regression RBF neural network and the introduction of least squares support vector machines (SVM) algorithm ,by BP, RBF, generalized regression RBF and least squares support vector machine flotation refined coal ash soft measuring modeling comparison, in the circumstances of using small sample ,the model accuracy and generalization ability of the least squares support vector machine (SVM) which is based on statistics theory of learning can be well verified. It provide the reliable basis for the flotation process refined coal ash soft survey modeling which used the least squares support vector machines.

Keywords: Flotation cleaned coal ash, least squares support vector machine, generalized regression RBF neural network, Generalization ability.

1 Introduction

More and more attention are paid to artificial neural network and support vector machine in soft sensor modeling, Many scholars conducted fruitful research [2], In the process of slime flotation clean coal ash real-time measuring is becoming the key technology of flotation automation solutions [1], Adopting soft measuring modeling method is important method to solve the real-time detection of the clean coal. In this paper the experimental data of Zhangcun Coal Preparation Plant is put as the simulation data, using generalized regression RBF neural network model and the LS - SVM modeling method for the flotation clean coal ash soft sensor modeling simulation comparisons. Advantages and disadvantages of LS-SVM modeling and neural network are discussed in prediction accuracy and generalization ability.

2 Algorithm Description

This thesis algorithm involves BP, RBF, generalized regression RBF three neural network model and LS – SVM. Because of space and for BP, RBF neural networks are familiar; this paper introduced only generalized regression RBF neural network and the LS – SVM two algorithm principles.

2.1 Generalized Regression Network Algorithm Principle

Generalized regression neural network is radial basis function neural network, Its theory is based on nonlinear nuclear regression analysis, The dependent variable y relative to the independent variable x of regression analysis is actually calculating the maximum probability value the y , A random variable x and y the joint probability density function $f(x, y)$, Obser -vations of known value x X , then the return of y relative to X . Namely conditions mean:

$$\bar{Y} = E[y|X] = \frac{\int_{-\infty}^{+\infty} yf(X, y)dy}{\int_{-\infty}^{+\infty} f(X, y)dy} \tag{1}$$

For unknown probability density function $f(x, y)$, x and y can be observed sample obtained by the non-parametric estimation

$$\bar{f}(X, Y) = \frac{1}{(2\pi)^{\frac{(p+1)}{2}} \delta^{p+1} n} \times \sum_{i=1}^n \exp\left[-\frac{(X-X_i)^T(X-X_i)}{2\delta^2}\right] \exp\left[-\frac{(Y-Y_i)}{2\delta^2}\right] \tag{2}$$

Among them: X_i, Y_i is a random variable x and y values of the sample observations; δ is smooth factors; N for sample size; P is random variable x dimension.

With $\bar{f}(x, y)$ instead of $f(x, y)$, generation into type, and exchange integral and summation order, available:

$$\bar{Y}(X) = \frac{\sum_{i=1}^n \exp\left[-\frac{(x-x_i)^T(x-x_i)}{2\delta^2}\right] \int_{-\infty}^{+\infty} y \exp\left[-\frac{(y-Y_i)^2}{2\delta^2}\right] dy}{\sum_{i=1}^n \exp\left[-\frac{(x-x_i)^T(x-x_i)}{2\delta^2}\right] \int_{-\infty}^{+\infty} \exp\left[-\frac{(y-Y_i)^2}{2\delta^2}\right] dy} \tag{3}$$

For the integral items, Use properties $\int_{-\infty}^{+\infty} x \exp(-x^2) dx = 0$ for reduction:

$$\bar{Y}(X) = \frac{\sum_{i=1}^n Y_i \exp\left[-\frac{(x-x_i)^T(x-x_i)}{2\delta^2}\right]}{\sum_{i=1}^n \exp\left[-\frac{(x-x_i)^T(x-x_i)}{2\delta^2}\right]} \tag{4}$$

In type, Estimate for all samples $\bar{Y}(X)$ weighted average of observations Y_i , each observation Y_i weighting factor is Euclid distance (between the corresponding sample X_i and X) square index. When smooth factor δ value very large $\frac{(x-x_i)^T(x-x_i)}{2\delta^2}$ tend to zero, Estimate $\bar{Y}(X)$ is equal to all the dependent variable mean template. Conversely, when smooth factors tend to zero, Estimate $\bar{Y}(X)$ and the training sample are very close.

When the forecast point is to be included in the training sample set, Equation find the predicted dependent variable and the sample is very close to the corresponding dependent variable. But when it comes to sample the point, not included may forecast effect will be very poor. This phenomenon is called over fitting. When smooth factor

value moderate, When ask estimate $\bar{Y}(X)$, All the training samples of the dependent variable were into consideration. And the dependent variable which is corresponding to sample points which is closer to the point estimated was weighted more.

2.2 Least Squares Support Vector Machine Modeling Principle [3]

LS-SVM is based on the objective function of the standard SVM increased error squares items. Namely The main difference between least squares support vector machine, (LS-SVM), and support vector machine (SVM) is that the second square of error will be as loss function , not sensitive loss function as a loss function , This allows will change inequality constraints into a equality constraint conditions. For nonlinear systems, the nonlinear regression function is regarded:

$$f(x) = w^T \phi(x) + b \tag{5}$$

Among them: $x \in R^n, y \in R$, Nonlinear function $\phi(\cdot) R^n \rightarrow R^{n_k}$

The input space data through the mapping function is mapped to high dimensional feature space. For a given training data set $\{x_k, y_k\}_{k=1}^N$, LS-SVM can define the following optimization problems, γ for regularization parameters

$$\min_{w,b,e} J(w,e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{k=1}^N e_k^2 \tag{6}$$

s.t. $y_k = w^T \cdot \phi(x_k) + b + e_k, k=1, \dots, n$

Given “Lagrange” function for:

$$L = J - \sum_{k=1}^N a_k [w^T \cdot \phi(x_k) + b + e_k - y_k] \tag{7}$$

In the type Lagrange is a multiplier.

According to a conclusion, that is: KTT conditions again

$$\begin{aligned} \frac{\partial L}{\partial w} = 0 &\rightarrow w = \sum_{k=1}^N a_k \phi(x_k) \\ \frac{\partial L}{\partial b} = 0 &\rightarrow \sum_{k=1}^N a_k = 0 \\ \frac{\partial L}{\partial e_k} = 0 &\rightarrow a_k = \gamma e_k \\ \frac{\partial L}{\partial a_k} = 0 &\rightarrow w^T \cdot \phi(x_k) + b + e_k - y_k = 0 \end{aligned} \tag{8}$$

For $k = 1 \dots N$, expunction the w and e , gets below equation:

$$\begin{bmatrix} 0 & 1^T \\ 1 & K + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ Y \end{bmatrix} \tag{9}$$

Type:

$1=[1, \dots, 1]^T, Y=[y_1, \dots, y_N]^T, a=[a_1, \dots, a_N]^T$. K as a phalanx, $K_{ij} = \phi(x_i)^T \phi(x_j) = k(x_i, x_j)$, $K(\dots)$ is a Kernel function. Further by using least-square method to solve coefficient with a and b , finally the output which SVM model predict is:

$$y(x) = \sum_{k=1}^N a_k \phi(x)^T \phi(x_k) + b = \sum_{k=1}^N a_k k(x, x_k) + b \tag{10}$$

2.3 The Performance of the Soft Sensor Modeling of Evaluation Indexes

For predictive model, the key to good results in order to obtain the predicted value. Evaluation is used to evaluate the merits of the prediction model; this paper adopts RMS error to measure system error. If the RMS relative error numerical small, it means that the prediction is near the real value. Prediction effect is good. Otherwise prediction effect is poor. RMS error is defined as:

$$MSE = \frac{1}{l-1} \left(\sum_{i=1}^l |y_i - y_i^*|^2 \right) \tag{11}$$

3 Neural Network and Support Vector Machines Contrast

Comprehensive domestic and foreign theory study and practical application to generalized regression neural network. Generalized regression neural network prediction model has the following advantages:

- (1) Good misalignment mapping ability
- (2) Modeling need less sample size
- (3) The artificial determination's parameter is few

The generalized recurrent nerves network's learning algorithm does not adjust between neuron's connection weight in the training process. The network study is decided completely by the data sample, only needs to determine the smooth parameter. Especially the generalized recurrent nerves network does not need to carry on the artificial determination to the network architecture parameter which has the tremendous influence on the model predictive ability such as Transfer function, implicit strata neuron integer and so on. This characteristic had decided the generalized recurrent nerves network forecast model can maximum limit avoid the artificial subjective hypothesis to forecasting result influence.

Simulation results show that: The test set mean-square error $MSE = 0.019972$, The sample set mean-square error $MSE = 0.024332$. This is better than LS-SVM modeling ($\text{gam} = 3306$ $\text{sig2} = 14.9$), this proof generalized regression RBF neural network generalization ability is not poorer than the LS - SVM support vector machine.

By reducing the training set samples and increasing the testing set samples, then observe training set smaller cases generalized regression RBF neural network and the LS-SVM forecasting precision. By selecting : Zhongcun flotation process data 50% training set - 50% test set, 35% training set - 65% respectively with generalized regression test set RBF neural network and the LS - SVM modeling do contrast.

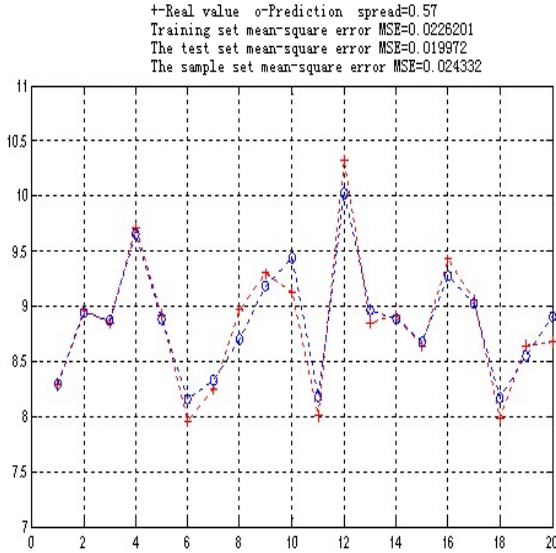


Fig. 1. Generalized Regression RBF Neural Network Simulation

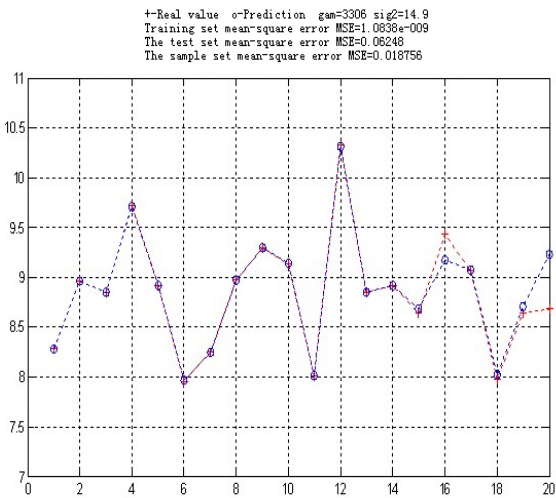


Fig. 2. LS-SVM Simulation

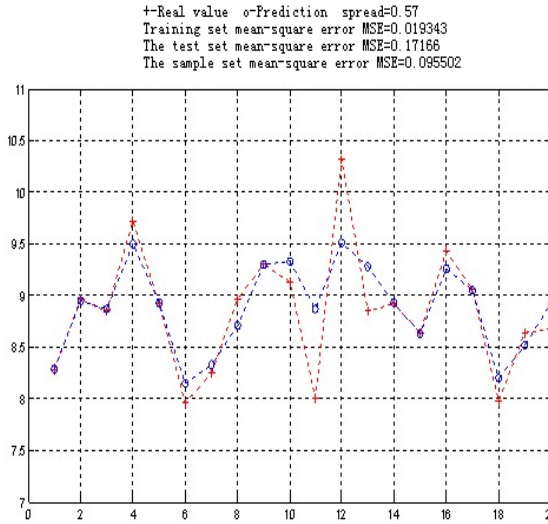


Fig. 3. Generalized Regression RBF Neural Network Simulation (50% training set - 50% test set)

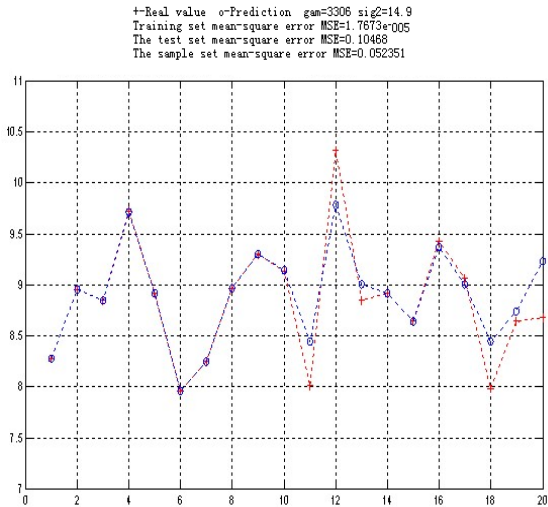


Fig. 4. LS-SVM Simulation (50% training set - 50% test set)

Two reduced training regulations' opposite experiment following chart shows:

LS-SVM to test sets mean-square error and the whole sample mean square error is much smaller than generalized regression neural network. It can be seen that when the training sample is few, LS-SVM situation also has smaller forecasting errors than generalized regression neural network.

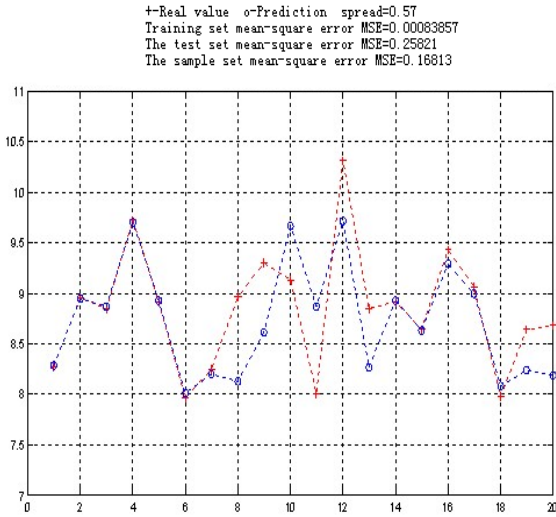


Fig. 5. Generalized Regression RBF Neural Network Simulation (35% training set - 65% test set)

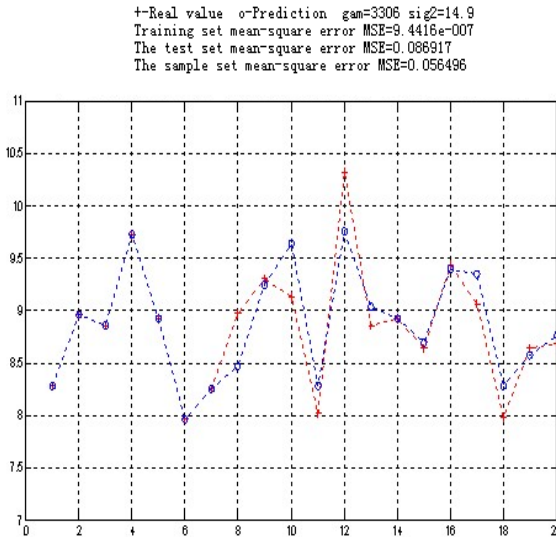


Fig. 6. LS-SVM Simulation (35% training set - 65% test set)

LS-SVM and other three kind of modeling method contrast result as shown in Table 1, From Table 1 can be seen, in the circumstances of Reducing learning samples, BP network, RBF network, generalized regression function approximation and return in the RBF all have good performance . But in the case of reducing study sample, generalization ability of LS - SVM is the best.

Considering that function returns and generalization performance having coequal and important requirements to the online clean coal ash prediction accuracy, The LS-SVM soft survey modeling is more suitable to take the online refined coal ash forecast algorithm use.

Table 1. Different Models MSE Comparison

Modeling type	70% training set - 30% test set	50% training set - 50% test set	35% training set - 65% test set
BP	0.033365	0.18354	0.21276
RBF	0.10088	0.18451	0.20343
Generalized RBF	0.019972	0.17166	0.25821
LS-SVM	0.06248	0.10468	0.086917

4 Conclusion

Through four network model prediction performance simulation comparison, it can be seen that: in the case of reducing study sample BP network, RBF network, generalized regression RBF, generalization ability is clearly worse than LS - SVM, Based on small sample conditions and statistical learning theory basis, LS - SVM in two aspects of precision and generalization performance achieve good compromise. Simulation shows that The LS-SVM soft survey modeling is more suitable to take the refined coal ash soft survey modeling algorithm use.

Acknowledgments. This work was supported by Natural Science Foundation of ShanXi Province (Item Number: 2009011034).

References

1. Al-thyabat, S.: On the optimization of froth flotation by the use of an artificial neural network. *J. China Univ. Mining&Technol.* 18, 418–426 (2008)
2. Schlang, M., Lang, B., Poppe, T., et al.: Current and future development in neural computation in steel processing. *Control Engineering Practice* 9, 975–986 (2001)
3. Suykens, J.A.K., Vandewalle, J., De Moor, B.: Optimal Control by Least Squares Support Vector Machines. *Neural Networks* 14, 23–35 (2001)
4. Jamsa-Jounela, S.-L., Vermasvuori, M., Enden, P., et al.: A process monitoring system based on the Kohonen self-organizing. *Control Engineering Practice* 11, 83–92 (2003)

The Research on Image Retrieval Based on Combined Multi-Features and Relevance Feedback

Shu-Juan Zhang

School of Electronics and Computer Science Technology, North University of China, Taiyuan,
030051, China
wqm1712@126.com

Abstract. The color feature is demonstrated by the algorithm of the improved histogram. The texture feature is extracted by Gabor filters. On the basis of above contents, the article studies a method for image retrieval using combined color feature and texture feature. Then by studying the theory of support vector machines, the algorithm of the SVM relevance feedback is introduced. The results of experiments show that combined feature extraction and relevance feedback algorithm has better retrieval performance and the results can be obtained to better meet the need of users.

Keywords: Image retrieval, Color feature, Texture feature, Support vector machines, Relevance feedback.

1 Introduction

With the popularity of the Internet and the rapid development of multimedia technology, the application of image information becomes more extensive. In order to facilitate the users to find quickly and accurately the contents needed from a large number of image information, the traditional method of text-based image retrieval has been unable to meet actual needs. In order to overcome the limitations of traditional retrieval methods, content-based image retrieval technology came into being.

Content-based image retrieval used mainly the underlying features of image color, texture, shape and spatial location relationship. Color has a little dependence on the size and direction of the image, so using color feature of images to retrieve image is the essential method of content-based image retrieval technology. But color feature is sensitive to brightness and noise. The texture feature have a stronger resistance to the noise, but also to a certain extent, the texture feature can describe the spatial information of the image. Therefore, to use the method of combining color features with texture features to retrieve images will made up the deficiencies of using only a single features to retrieve images. On the other hand in order to solve semantic gap well between low-level visual features and high-level human language, has introduced the SVM relevance feedback. Through continuous interaction with the system, the effectiveness of retrieval results has enhanced.

2 Using Color Feature to Retrieve Image

In order to use correctly the color feature to retrieve images and the results of image retrieval more in line with the human visual sensory, first of all, it is necessary to create the color space model. In practice, there are many the color space models mostly used, but to select the color model have a great impact on the retrieval results. The HSV color space model can better meet the perceived characteristics of the human's eyes. At the same time, the HSV can to converse easily with RGB color space model. Therefore, HSV color space model is used in the article. HSV space will be quantified non-interval, for each component of the quantized, to join them into one-dimensional feature vector, that is, $G = HQSQV + SQV + V$, among which, QS, QV are quantitative series of components. Normally, QS = 4, QV = 4, then $G = 16H + 4S + V$. G is in the range of [0,255]. Therefore, to calculate the G will get the one-dimensional histogram vector of 256 handle.

The retrieval method of a global color histogram is simple and quick, it only contains color information, but that do not reflect the overall relative position, which is bound to result in retrieval errors. In this regard, the article uses a modified color histogram retrieval algorithm, the whole image is divided into $m \times n$ sub-blocks. According to the image background and the spatial relations and the importance of the central object, it is necessary to increase weight for some sub-blocks. The intermediate sub-block usually has larger weight in the image, but the edge of the sub-block has smaller weight. The retrieval algorithm is described as follows:

(1)The each image file is divided into $(m \times n)$ sub-image I_{ij} in the target image library, and extract the features to form feature library.

(2)Calculating local color histogram of the each sub-image Q_{ij} ,then calculating the similarity measure about the sub-image Q_{ij} and I_{ij} ,that is

$$S_{ij}(Q_{ij}, I_{ij}) = \frac{1}{N} \sum_{k=1}^N [1 - \frac{|q_k - i_k|}{\max(q_k - i_k)}], (i = 1, 2, \dots, m; j = 1, 2, \dots, n) \tag{1}$$

N mean the color series, q_k mean the frequency of k-color in the sub-image Q_{ij} , i_k mean the frequency of k-color in the sub-image I_{ij} .

(3)Considering the weight of the sub-block, that is W_{ij} , then calculating the similarity measure about the Q and I,that is

$$S(Q, I) = \sum_{i=1}^m \sum_{j=1}^n W_{ij} \times S_{ij}(Q_{ij}, I_{ij}) \tag{2}$$

(4)Given the threshold, that is $\delta(0 < \delta < 1)$, the retrieval results is the image of $S(Q, I) > \delta$.

3 Using Texture Feature to Retrieve Image

Texture feature is a kind of visual features which reflect homogeneous phenomenon of image, which does not rely on color or value, which has been widely applied in content-based image retrieval. The methods of texture description include the Statistics, the Spectrum and the Construction. The article uses the Spectrum method,

and uses Gabor filters to extract texture feature of image. The two-dimensional Gabor transform used in texture analysis is as follows:

$$g(x, y) = \frac{1}{2\pi\delta_x\delta_y} \exp[-\frac{1}{2}(\frac{x^2}{\delta_x^2} + \frac{y^2}{\delta_y^2}) + 2\pi jWx] \tag{3}$$

W is the complex modulation frequency of Gauss function. g(x, y) for the mother-wavelet, by appropriate scaling and rotation transformation to g(x, y), that is, by changing the following values of m and n, then a group of self-similar filter banks is gotten, which known as the Gabor wavelet, in which direction and scale are different.

$$g_{mn}(x, y) = a^{-m} g(x', y') \quad a > 1, m, n \in \mathbb{Z}$$

$$x' = a^{-m}(x \cos \theta + y \sin \theta) \tag{4}$$

$$y' = a^{-m}(-x \sin \theta + y \cos \theta) \quad \theta = n\pi/N$$

m=0,1,... M-1, n=0,1,... N-1, M is the number of scale, and N is the number of direction. Given an image I(x,y), Gabor wavelet is transformed as follows:

$$W_{mn} = \iint I(x, y) g_{mn}(x - x_1, y - y_1) dx_1 dy_1 \tag{5}$$

Seeking the mean and the variance as follows:

$$\mu_{mn} = \iint |W_{mn}(x, y)| dx dy$$

$$\delta_{mn} = (\iint (W_{mn}(x, y) - \mu_{mn})^2 dx dy)^{1/2} \tag{6}$$

to obtain the eigenvector as follows:

$$f = [\mu_{00}, \delta_{00}, \mu_{01}, \delta_{01}, \dots, \mu_{mn}, \delta_{mn}] \tag{7}$$

4 Combined Multi-features Retrieval

Since the physical meaning of the color feature eigenvector and texture feature eigenvector is different, therefore, different eigenvector is carried on external normalized, which ensure that different feature vectors in the same position during the similarity calculation and the same role in the retrieval. Eigenvector external normalized uses the method of Gaussian normalization. The method is as follows:

(1) Assuming there are M images, calculating the similar distance between eigenvector F_X of images X and eigenvector F_Y of images Y, that is, D(X, Y) = dis(F_X, F_Y).

(2) Obtained by the above formula to calculate the distance values of the number of M(M-1)/2, then calculating the mean μ and the standard deviation δ of the values.

(3) To retrieve image Q, calculating the similar distance between image Q and each image, the similar distance are described as D_{1Q}, D_{2Q}, ..., D_{MQ}, and which has a linear transformation, as follows:

$$D_{IQ}' = (1 + (D_{IQ} - \mu) / 3\delta) / 2 \tag{8}$$

When using combined color feature and texture feature to retrieve images, users can set the weight coefficient of the color eigenvector and texture eigenvector according to the actual situation, so it is able to meet the needs of different retrieval.

The article gives the weight coefficient of color eigenvector and texture eigenvector: $\omega_1 = \omega_2 = 0.5$, and then proceeded to sum about linear weighted and calculate the similar distance between image Q and each image in database.

5 SVM-Based Relevance Feedback

5.1 Support Vector Machine Model

SVM is developed from the optimal classification surface in the linear separable case. It is the learning algorithm based on the principle of structural risk minimization. The optimal classification surface is constructed in high dimensional space by in the original space or projection. The given training samples belonging to two categories are separated. The distance that two samples from the hyperplane is maximum, which is the basis of hyperplane constructed. Assumption Sample set of linear separable $(\vec{x}_i, y_i), i = 1, 2, \dots, n; \vec{x} \in R^d, y \in \{+1, -1\}$ is category label. The general form of linear discriminant function in the d-dimensional space is $g(x) = \vec{w} \cdot \vec{x} + b$. The equation of the classification surface is $\vec{w} \cdot \vec{x} + b = 0$. The discriminant function is normalized. All samples are to meet $|g(x)| \geq 1$. The classification interval is equal to $2 / \|\vec{w}\|$. It is requirements for all samples correctly classified, which is to meet

$$y_i(\vec{w} \cdot \vec{x}_i + b) - 1 \geq 0, i = 1, 2, \dots, n \tag{9}$$

the optimal classification surface makes classification interval reach to the maximum under the conditions of the expression (9), that is $\|\vec{w}\|^2$ to attain the minimum. Solving the optimal classification surface can be expressed as the following optimization problem, that is to attain the minimum of the expression (10).

$$\Phi(\vec{w}) = \left\| \vec{w} \right\|^2 / 2 = (\vec{w} \cdot \vec{w}) / 2 \tag{10}$$

The Lagrangian function is used to solve this optimization problem, which obtains optimal solution α_i^* and $\vec{w} = \sum_{i=1}^N \alpha_i^* y_i \vec{x}_i$. The classification function is

$$f(\vec{x}) = \text{sign}\{(\vec{w} \cdot \vec{x}) + b^*\} = \text{sign}\left\{\sum_{i=1}^N \alpha_i^* y_i (\vec{x}_i \cdot \vec{x} + b^*)\right\} \tag{11}$$

In the case of linear inseparable kernel function K is used, which makes dot product operations become $K(\vec{x}_i, \vec{x}_j) = \phi(\vec{x}_i) \cdot \phi(\vec{x}_j)$, therefore the classification function become

$$f(\vec{x}) = \text{sign}\left\{\sum_{i=1}^N \alpha_i^* y_i K(\vec{x}_i, \vec{x}) + b^*\right\} \tag{12}$$

5.2 SVM-Based Relevance Feedback Algorithm

In the retrieval process the user evaluates and marks the retrieval results. Pointed out that the results of which are associated with the query image, which is not relevant. Then these samples were related and not related to SVM learning, which gets a SVM classifier that represents the retrieval goals of the users. All images are classified by the SVM classifier. The class is divided into positive images, calculated the distance of relative to the classification surface of each image. The image that is farther away from the classification surface is closer the query image. According to the distance, sort descending again and obtain results. The algorithm is described as follows:

(1) According to the image given the initial retrieval is carried on.

(2) The user mark and retrieve images related to the objectives, obtain related atlas I_1^+ and irrelevant atlas I_1^- in the feedback results, which is used to update the related sets I^+ and irrelevant sets I^- .

$$\begin{aligned} I^+ &= (I^+ \cup I_1^+) - I_1^- \\ I^- &= (I^- \cup I_1^-) - I_1^+ \end{aligned} \quad (13)$$

(3) Constructed training set (x_i, y_i)

$$x_i \in I^+ \cup I^-, y_i = \begin{cases} +1, & x_i \in I^+ \\ -1, & x_i \in I^- \end{cases} \quad (14)$$

(4) Training samples for learning by SVM and constructed classification

$$f(\vec{x}) = \text{sign} \left\{ \sum_{i=1}^N \alpha_i^* y_i K(\vec{x}_i, \vec{x}) + b^* \right\} \quad (15)$$

(5) The class is divided into positive images, calculated the distance of relative to the classification surface of all sub-image about each image. The maximum is taken as the distance between the image and the classification surface.

(6) Calculated from the above question and obtained the value of all the images, according to descending order, return results.

(7) If the user satisfaction feedback process is stopped, otherwise the retrieval results and the classification results of the user are added in the total sample set. According to sample set, to retrieve again until the retrieval results meet user requirements.

6 Experiments and Results

The 2000 images were chosen as the test database, including flowers, cars, animals, landscape, mountain, building, wood and lakes. The various images are more than 100. The article used the test library and designed the following experiments, including using color feature to retrieve image, using texture feature to retrieve image, carrying on combined retrieval by setting the weight coefficients of two eigenvectors, introduction of SVM-based relevance feedback. In comparing different methods of retrieval performance, the Precision and the recall is used as the evaluation criteria of the retrieval results. Precision ratio refers to the system returns the number of relevant

images in proportion to the number of returned images of all in the course of a retrieval. Recall ratio refers to the system returns the number of relevant images in the retrieval results in proportion to the number of all relevant images in the image test database. When doing the experiments, five types of images were chosen from the image test database. The images from each type were randomly chosen 10 as the retrieval image, generated 50 times the retrieval processes. The average precision ratio and average recall ratio about retrieval results were respectively calculated for each type of image. As shown in Table 1 (The data in the table is described by percentage).In the SVM-based relevance feedback experiments, the relation of feedback frequency and the average precision ratio is reflected in Figure 1.

Table 1. The average precision ratio and the average recall ratio by using different feature

Image type	Average precision ratio			Average recall ratio		
	Color	Texture	Combined	Color	Texture	Combined
flowers	56.5	45.1	63.2	57.1	47.5	64.3
cars	51.3	48.9	58.1	52.5	48.2	56.8
animals	47.6	41.8	53.7	48.0	38.8	46.8
mountain	40.2	45.6	56.3	38.2	43.7	52.1
wood	50.5	52.7	60.8	51.9	51.3	59.2

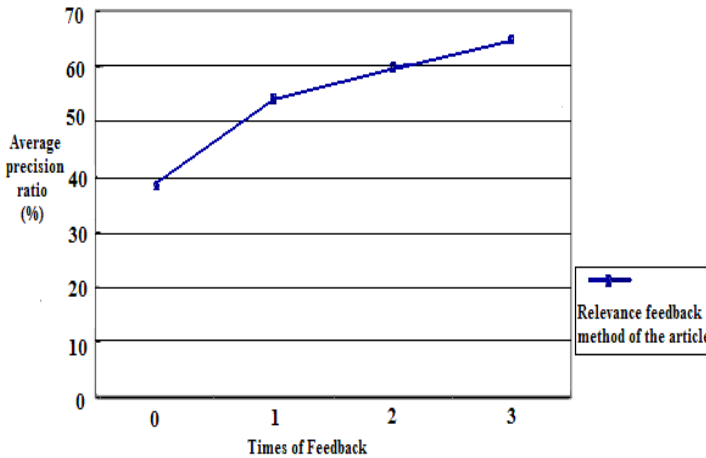


Fig. 1. Relevance feedback experiments

The results can be seen that combined multi-features retrieval has higher efficiency than a single feature's retrieval. With the increase in the number of feedback, the relevance feedback methods used in the article has better precision ratio.

7 Conclusion

Since image retrieval based on single feature has some limitations, the retrieval algorithm for each feature fully dose not reflect the information contained in the image. Therefore, multiple features in the image are combined. The image retrieval result is improved to some extent and realize the function of relevance feedback according to SVM-based relevance feedback algorithm. So that more similar images are retrieved, and retrieval results can be more close to the user's requirements.

References

1. Tao, D.C., et al.: Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28, 1088–1099 (2006)
2. Platan Iotis, K.N., Venetsanopoulos, A.N.: *Color Image Processing and Applications*. Springer, Heidelberg (2000)
3. Simona, E.G., Petkov, N., Kruizinga, P.: Comparison of Texture Features Based on Gabor Filters. *IEEE Transactions on Image Processing* 11, 1160–1167 (2002)
4. Dong, W., Zhou, M., Geng, G.: Image retrieval technique based on combined features. *Compute Applications and Software* 22, 34–36 (2005)
5. Finlavson, G., Schettine, R.: Special issue: color for image indexing and retrieval. *Computer Vision and Image Understanding* 94, 1–268 (2004)
6. Su, Z., Zhang, H., Ma, S.: Relevant feedback using a Bayesian classifier in content- based image retrieval. In: Yeung, M.M. (ed.) *Proceedings of the SPIE Storage and Retrieval for Media Databases*, vol. 4315, pp. 97–106. SPIE Press, San Jose (2001)
7. Howarth, P., Ruger, S.M.: Evaluation of Texture Features for Content-Based Image Retrieval. In: *Third International Conference*, pp. 326–334 (2004)
8. Razniewski, S., Strzelecki, M.: Evaluation of texture features based on mutual information. In: *Image and Signal Processing and Analysis*, September 15-17, pp. 233–238 (2005)
9. Zhang, H., Ou, Z., Li, G.: An image retrieval method on color primitive co-occurrence matrix. In: Jiao, L., Wang, L., Gao, X.-b., Liu, J., Wu, F. (eds.) *ICNC 2006*. LNCS, vol. 4222, pp. 526–529. Springer, Heidelberg (2006)
10. Kokare, M., Biswas, P.K., Chatterji, B.N.: Texture Image Retrieval Using New Rotated Complex Wavelet Filters. *IEEE Trans. on Systems, Man and Cybernetics* 35, 1168–1178 (2005)
11. Rui, M., Cheng, H.D.: Effective image retrieval using dominant color descriptor and fuzzy support vector machine. *Pattern Recognition* 42, 147–157 (2009)
12. Wan, H.L., Chowdhury, M.U., Hu, H., Shi, Z.Z.: Texture feature and its application in CBIR. *Journal of Computer-Aided Design&Computer Graphics* 15, 195–199 (2003)

Application of BP Neural Network in Drug Supervision Code Grade Evaluation

Jing Zhao and Ye-li Li*

Information and Mechanical Engineering Institute Beijing Institute of Graphic
Communication Beijing, China
liy1@bigc.edu.cn

Abstract. In order to implement the policy about strengthen drug safety supervision and management, National regulations the grade of drug supervision code must be above grade C [2]. In this paper, mainly selected a barcode optical parameters, in order to objective evaluate each index of the drug supervision code, adopted BP artificial neural network in neural network toolbox of MATLAB, used the data which measured from Barcode reflectance curve to establish BP neural network model of BP. The result indicates BP artificial neural network model can assess code every index and grade. The result also indicates that at the same time the BP artificial neural network in MATLAB neural network toolbox can be applied to the grade evaluation of the drug supervision code.

Keywords: BP neural network, drug supervision code, neurons, Optical parameters.

1 Introduction

In order to implement the policy about strengthen drug safety supervision and management, accelerate the establishment of a key drug safety traceability system, strength drug quality safety supervising, ensure public using the drug safely. State food and drug administration decisions, on special drug monitoring information network basis, further strengthen drug electronic supervision, perfected drug identification system, establish a nationwide unified supervision and administration of pharmaceutical electronic network, classification of drugs enforcement partial to electronic supervision.

Drug packaging supervision code testing involves many aspects: decoding the correctness of detection, optical characteristics parameters testing, can decode degree detection; Optical characteristics parameters including: Minimum reflectivity (R_{\min}), symbols contrast (SC), minimum edge contrast (EC_{\min}), modulation ration (MOD) and defect degrees (Defects). These five parameters from several aspects the barcode symbols and comprehensive evaluation on the optical properties of the reflectivity

* Corresponding author.

including color mates of bar (space), the contrast of adjacent bar (space), the influence of contrast by ink spread and the influence of local reflectivity by stripping. Each measure is to find out the data which is obtained from the barcode reflectance curve and calculated according to the formula the value of each parameter, thus determine their level.

In this paper, using many index to assess drug packaging supervision code level. Research the relationship about barcode reflectivity with the parameters, find related characteristics and training artificial neural network, is the key to determine the barcode level which based on the calculated data processing technology. Artificial neural network simulation of human thinking mode, advantage is based on data, through learning and training to master the change rule of internal system, do not need to use mathematical equation expressing the relationship between input and output, in the course of modeling avoid the effect of artificial factor. Not only a high degree of automation and the resolution of the evaluation results are satisfactory. It has a good application prospect.

2 Introduction of Bar Code Level Detection

Barcode symbol is a logistics information carrier of the business sector and other areas, in such objects in the flow of each chain, the computer will identify through barcode symbols which adhere to of the object, realize the information collection work of information system. Barcode detection is an important means which can ensure the barcode throughout the supply chain can be conversational right. Diagram vertical axis represents the value of reflectivity, horizontal axis represents length direction of position. High reflectivity area is space, low reflectivity area is bar. The two sides of the high reflectivity area were left, right blank area. The important characteristics of barcode reflectance curve can be obtained by manual graphics analysis or automated numerical analysis to identify.

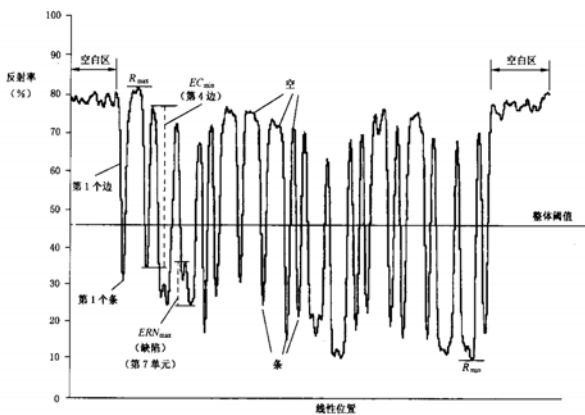


Fig. 1. Reflectance curve

2.1 Barcode Testing Parameters

According to GB/ T14258-2003 stipulated, the parameters of the need to detect are: decode correctness, minimum reflectivity, symbols contrast, minimum edge contrast, modulation ration, defect degrees, can decode degrees, symbol standard and use standard of scanning reflectance curve additional requirements.

Optical characteristics parameters testing: minimum reflectivity (R_{\min}), symbols contrast (SC), minimum edge contrast (EC_{\min}), modulation ration (MOD), defect degrees (Defects), these five parameters from several aspects the barcode symbols and comprehensive evaluation on the optical properties of the reflectivity including color mates of bar (space), the contrast of adjacent bar (space), the influence of contrast by ink spread and the influence of local reflectivity by stripping. Each measure should be found the minimum reflectivity and the maximum reflectivity from the barcode reflectance curve, use the equation (1) calculate barcode symbols contrast SC; Find out the adjacent unit (contain blank area) space reflectivity (R_s) and the bar reflectivity (R_b), use the equation (2) calculation of edge contrast (EC); Find out minimum namely minimum edge contrast (EC_{\min}); use the equation (3) calculation modulation ration (MOD); Defect degrees is units and blank area of inhomogeneous degree of reflectance curve, represent by the value for the difference with lowest reflectivity and peak reflectivity. If a unit without peak or empty, so ERN is 0, so use the maximum ERN to represent the max inhomogeneous degree of reflectance curve (ERN_{\max}), according to the formula (4) get defect degrees.

$$SC = R_{\max} - R_{\min} \quad (1)$$

$$EC = R_s - R_b \quad (2)$$

$$MOD = EC_{\min} / SC \quad (3)$$

$$Defects = ERN_{\max} / SC \quad (4)$$

2.2 Optical Parameters Level Determination

Through the above form, we can draw every optical characteristics parameters level uncertainty range. GB / T14258-2003 specified in the self-identify one solution is to get level of all parameters inside the level of minimum as barcode level series, therefore barcode optical properties determine the parameters level is also will all the optical properties of the minimum level of the parameters of the optical properties parameters as bar code level.

Grade	R_{min}	SC	EC_{min}	MOD	Defects
4	$R_{min} \leq 0.5R_{max}$	$SC \geq 70\%$	$EC_{min} \geq 15\%$	$MOD \geq 0.70$	$Defects \leq 0.15$
3		$55\% \leq SC < 70\%$		$0.60 \leq MOD < 0.70$	$0.15 < Defects \leq 0.20$
2		$40\% \leq SC < 55\%$		$0.50 \leq MOD < 0.60$	$0.20 < Defects \leq 0.25$
1		$20\% \leq SC < 40\%$		$0.40 \leq MOD < 0.50$	$0.25 < Defects \leq 0.30$
0	$R_{min} > 0.5R_{max}$	$SC < 20\%$	$EC_{min} < 15\%$	$MOD < 0.40$	$Defects > 0.30$

Fig. 2. Optical parameters rating table

3 BP Network Structure of Optical Parameters Grades and the Training Process

Artificial neural network, by imitating animal’s neural network behavior, a kind of mathematical model distributed parallel information processing algorithm. This network rely on the complexity of the system, by adjusting the internal mutual connection of node relationship, thus reach the purposes of processing information. Artificial neural network has self-learning and adaptive ability. Analysis and masters the potential law between corresponding to the input and output data, finally, according to these laws with new input data to calculate output. Artificial neural network is by artificial neural as the basic unit constitution Neural network technology is one of the simplest abstraction and simulation to the human brain, a good tool to explore the mysteries of the human intelligence.

3.1 The Introduction of BP Neural Network Theory

In 1980s, David Hinton, Geoffrey Rumelhart and Williams independently given the clear indication of BP algorithm, solved the multilayer neural network learning problems. The BP neural network called the back-propagation neural network, widely used in pattern recognition, image processing, function approximation and optimization calculation, the optimal forecasting and adaptive control, etc.[3][4] BP network composed of two process which including the positive dissemination of information and error back propagation. The structure of the BP neural network model belongs of multilayer shape to network, contains input layer, hidden and output layer. Each neuron of input layer is responsible for receive the input information from outside, then passed to the neuron of middle; hidden is the internal information processing layer, be responsible for information transform, according to the demand for capacity information change, hidden can design for single hidden or more hidden structure; the last hidden transfer information to output layer, upon further processing, complete a learning process of positive spread, the output layer output information processing results to outside world. When output and when expected output discrepancies, enter back propagation phase. Error through the output layer gradient descent way correction weights of each layer, to hidden layer and input layer back propagation step by step. Information positive dissemination and error back propagation process, is the process of continuously adjust weights, is the process of

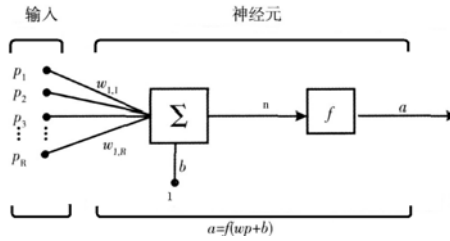


Fig. 3. BP nerve cell model

neural network learning training. As far as the network output error reduced to acceptable levels or Achieve predetermined number of learning so far.

3.2 Network Structure and Training of Optical Parameters Levels

The BP neural network of optical parameters grades for three layer structure, input layer with eight neurons, the number of neurons is through the training then singled out the best. The output layer including 5 neurons, and to identify parameters correspond to the output. Neural network to the five parameters for expected output vector is $T \{ y_1, y_2, \dots, y_5 \}$. Network training will use 100 known as the training sample set to bar neural network for training, network error Settings for 10^{-5} , training process adopts automatic stop, once network output error achieve predetermined error requirements, network training process will automatically stop. In this paper, three layer of network trained can make the network error drop to 10^{-5} , to achieve the desired accuracy. The graph is network training process with the change of iteration error.

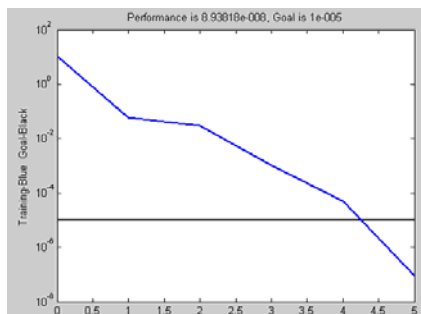


Fig. 4. Network error performance figure with trainlm

```

t1 =
    71.9981    10.0145    42.4793    58.9611    18.0127
y1 =
    0.8611    0.0002    0.4511    0.6800    0.1113
huanghe_t =
    72.0000    10.0000    42.5000    59.0000    18.0000
    
```

Fig. 5. Training results with trainlm

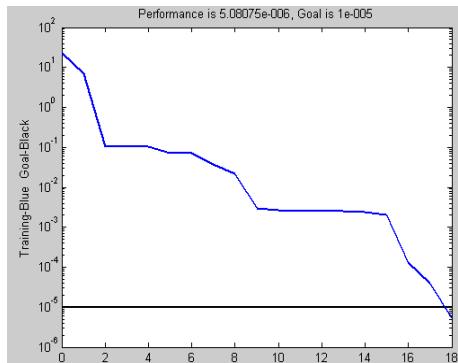


Fig. 6. Network error performance figure with trainbfg

```

t1 =
    72.0697    9.8921    42.2009    59.1511    18.0541
y1 =
    0.8621   -0.0015    0.4472    0.6827    0.1119
huanghe_t =
    72.0000    10.0000    42.5000    59.0000    18.0000
    
```

Fig. 7. Training results with trainbfg

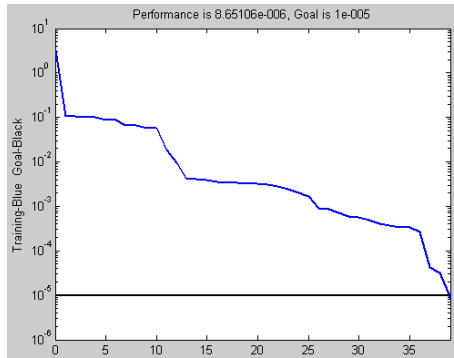


Fig. 8. Network error performance figure with trainscg

```
t1 =
    72.2206    10.1737    42.3319    58.6617    18.0518 ,
y1 =
    0.8642    0.0024    0.4491    0.6759    0.1118 ,
huanghe_t =
    72.0000    10.0000    42.5000    59.0000    18.0000 ,
```

Fig. 9. Training results with trainscg

3.3 The Analysis and Contrast of Results

First we will preprocess the input data of neural network. In this paper, we select three training function to train the network. If the train function selection trainlm, get the number of neurons in hidden layer for five, after five iterations, network have reached the expected error performance and the network has a very good simulation result. If the train function selection trainbfg, the number of neurons in hidden layer for seven, after eighteen iterations, network will reach the expected error performance, but, the result is not very good as previous. If the train function selection trainscg, the number of neurons in hidden layer for three, after thirty six iterations, network will reach the expected error performance. So, from three training function of the corresponding training results we can be seen that network training function selection trainlm, the simulation results are better.

4 The Conclusion

Artificial neural network has many advantages, for examples, self-organizing, self-learning and fault tolerance, Overcome traditional barcodes some unfavorable factors level method, And for a lot of data step type parallel processing, Greatly improved

barcode level identification accuracy, meanwhile shorten testing barcode optical characteristics parameters levels of time.[1].In this paper, Established BP neural network parameters model which is using to detect the barcode optical characteristics, improved BP network learning algorithm, further made barcode optical characteristics parameters level detection experiments. Experimental results show that, through the training, the BP network and the algorithm can avoid the local minimum, and the rate of convergence is relatively fast, can accurately detect the levels of barcode optical characteristic parameters. This research has reference value for other relevant parameters.

Acknowledgement. The authors would like to thank “Funding Project for Academic Human Resources Development in Institution of Higher Learning Under the Jurisdiction of Beijing Municipality(PXM2010_014223_095557)”and “Scientific Research Common Program of Beijing Municipal Commission of Education(Grant No. KM201010015003)”.

References

1. Kai, Z., Zhenglin, W.: Proficient in MATLAB neural network, pp. 102–103. Electronic Industry Press, Beijing (2010)
2. GB/T14258-2003 Bar code Symbols printing quality inspection
3. Haykin, S.: Neural network theory, vol. 1. Machinery Industry Press, Beijing (2004)
4. Zhihua, Z., Cungen, C.: Neural Network and Application, vol. 9. Tsinghua University Press, Beijing (2004)
5. Hagan, M.T., Demuth, H.B., Beale, M.H.: Neural network design. In: Dai Kui Translation, pp. 16–19. Machinery Industry Press, Beijing (2002)
6. Muvai, H.: Remote Sensing Image Analysis Using a Neural Network and Knowledge-based Processing. *International Journal of Remote Sensing* 18(4), 811–828 (1997)
7. Hagan, M.T., Menhaj, M.: Training Feed forward Networks with the Marquardt Algorithm. *IEEE Tran. on Neural Networks* 5(6), 989–993 (1994)

WordNet-Enhanced Dynamic Semantic Web Services Discovery

Li Chen, Zi-lin Song, Ying Zhang, and Zhuang Miao

Institute of Command Automation, PLA University of Science and Technology
Nanjing 210007, China
ivan.chen@163.com

Abstract. Web services are software systems designed to support interoperable machine-to-machine interaction over a network. There are mainly two problems in Web services discovery: how to describe service accurately and delicately to support precisely matchmaking and how to store, index and exchange the information of Web services in order to extend the scope of service discovery and guarantee the acceptable processing time of service discovery. In this paper we proposed a method for semantic Web services discovery using structured P2P technology. The service is published by the output concepts of service model in OWL-S. The service discovery is a two-steps task, service location through output concepts and service matchmaking through input concepts. Furthermore, we proposed the hashing technology based on synsets of WordNet to overcome the problem of lack of semantic information and only supporting exact searching in structured P2P networks.

Keywords: Web services, semantic Web, P2P, WordNet.

1 Introduction

Service-oriented computing is a computing paradigm based on the Service-Oriented Architecture (SOA)[1], which is an architectural style for building software applications out of services available in a network. SOA promotes loose coupling between software components so that they can be reused. Services are implementations of well-defined business functionality, and such services can be consumed by clients in different applications or business processes.

SOA uses the find-bind-execute paradigm whereby service providers register their services in a public registry. This registry is used by consumers to find services that match certain criteria. If the registry has such a service, it provides the consumer with a contract and an end point address for that service. The most widely used technology for realizing the SOA paradigm is Web services. Web services are software systems designed to support interoperable machine-to-machine interaction over a network. This interoperability is gained through a set of XML-based open standards such as the Web Services Description Language (WSDL) [2], the Simple Object Access Protocol (SOAP)[3], and the Universal Description, Discovery and Interaction (UDDI)[4].

These standards provide a common approach for defining, publishing, and using Web services.

With the development of Web services technology, the number of available Web services is growing day by day. How to find desirable services for the use's demand is then becoming the hot issue both in academy and industry. There are mainly two problems in Web services discovery. First problem is how to describe service accurately and delicately in order to support precisely matchmaking between Web services and use's demand; another problem is how to store, index and exchange the information of Web services in order to extend the scope of service discovery and guarantee the acceptable processing time of service discovery.

In traditional architecture of Web services, the description of Web services is stored in centralized registry. This kind of architecture assumes that every Web service coming on line advertises its existence and its capabilities/functionalities with the registry; and then every service requester contacts the registry to discover the most appropriate Web service and gather information about it. The Web service discovery is mostly the keyword-based or limited-property-based methods. It's insufficient to find the desirable services just based on keywords or limited properties. In order to overcome this problem, the technology of semantic Web is introduced to the Web services. With usage of ontology and logic reasoning systems, semantic Web services can be more machine-understandable and machine-processable. Furthermore, semantic Web services can describe Web services and use's demand more precisely and realize the logic matchmaking between them. The usage of semantics in centralized registry commendably solves the first problem we discussed above. OWL-S is the widely used model in the semantic Web services. The OWL-S proposes a semantics-based description of Web services. In OWL-S, each service is provided with a description consisting of three elements: the service profile (what the service does), service model (how the service works), and service grounding (how to access the service). In this paper, we commit to the use of OWL-S for describing Web services because of the enriched semantic capabilities it provides.

Semantic-enhanced centralized registries are effective since they guarantee discovery of services that have registered. On the other hand, the disadvantages in that case are that the existence of registries influences the overall performance very negatively and that any failure on the function of the registry would lead to a system failure (single point failure). Furthermore With the development of semantic Web services, the logic matchmaking is time consuming and more complicated than keyword-based matchmaking. It's ineffective to process semantic reasoning and matchmaking with large number of Web services in a registry (performance bottleneck). The expansibility and flexibility of service discovery become the crucial problem to be solved.

P2P technology is known to be a simple and powerful computing technology. A P2P system functions through the interaction between 'neighbors' thus avoiding central control and the single point of failure. This interaction is defined through the algorithms that enforce the topology and the roles of the participants. Each participant can

store/access other participant's data, if needed, with or without knowing its origin. If a participant fails, other participants can take over rebuilding the system. P2P technology also allows load distributing among participants. P2P networks can mainly be divided into three types according to the discovery/communication model: the centralized directory model, the flooded requests model, and the document routing model (usually called the distributed hash table model). The latter two have been gaining most attention from researchers due to their better flexibility and dependability than the first one.

Nowadays, the combination of P2P and semantic Web services for better service discovery and matchmaking has recently attracted the attention of different research communities. The methods of Web services discovery based on P2P mainly include several aspects as below.

Castro et al. [5] propose building a universal ring in a DHT P2P network to support service advertisement, service discovery, and code binding. These functions are based on three operations, namely the persistent store, the application-level multicast, and the distributed search. Every exposed service (a piece of code) has a code certificate to identify the correctness of the service (code binding). Since there may be many Web services with the same functionality and name globally, the code certificate will help the user find the correct service to invoke. It is essential for finding the correct service that the information used to generate the service key for discovery should be the same as the one used to generate the service key for advertisement. This is the problem of lack of semantic information and only supporting exact searching in structured P2P networks. If the name of service request isn't equal to the name of service advertisement, such that the result of hashing service request can't equal to the result of hashing service advertisement, the registry will return failure because none of service advertisements can match the service request.

METEOR-S Web Services Discovery Infrastructure (MWSDI, [6]) is a scalable publication and discovery environment involving semantic Web services and the P2P topology. In the MWSDI environment, there are many standard UDDI registries running in different Operator peers respectively. Every Operator peer exposes Operator Services to other peers. Operation Services are a set of services based on standard UDDI specification and enhanced by semantic technologies. The communication between peers relies on the P2P protocol. The user who needs to publish or discover services should call the Operator Services through a client peer, which is a transient interpreter between the user and MWSDI. To build the relationship between the Web service and the ontology, MWSDI uses WSDL and predefined tModels (the metadata structures in UDDI describing the common concept and understanding) to present taxonomies of inputs and outputs.

In order to overcome the disadvantage of the methods above, in this paper we proposed a method for semantic Web services discovery using structured P2P technology. The service is published by the output concepts of service model in OWL-S. The service discovery is a two-steps task, service location through output concepts and service matchmaking through input concepts. Structured P2P

network is the foundation of service publication and discovery. Traditional structured P2P network such as Chord lacks of semantic information and only supports keywords-based exact searching. In this paper we proposed the hashing technology based on synsets of WordNet to improve the precision of service discovery.

The rest of this paper is organized as follows. In section 2, we mainly introduce some relevant technologies such as OWL-S and a kind of structured P2P network Chord. Section 3 is the method for semantic Web service discovery using P2P technology. The last section is conclusions and our future works.

2 Relevant Technologies

2.1 OWL-S

Nowadays, most widely used semantic Web services model is OWL-S. OWL-S, apart from just being a language to describe services, defines an ontology structure for Web services as shown in figure 1.

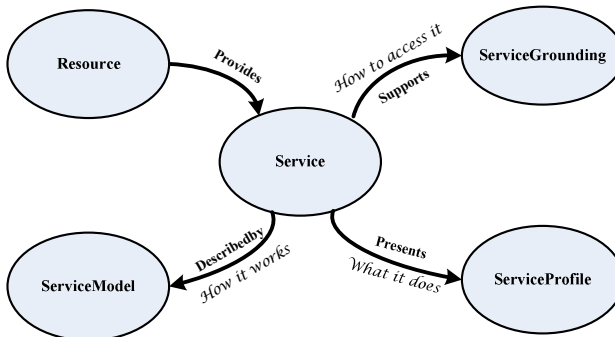


Fig. 1. Top level of OWL-S

The service profile is for publishing and discovering services, which explains what the service does (at an application level), what is accomplished by the service, limitations on service applicability and quality, and requirements that the service requester must satisfy to use it successfully. The process model is for describing the service's operation, which contains how it does it (at a computational level), the conditions under which particular outcomes will occur and the corresponding partial processes. This information can be further used for more accurate matching when searching. The service grounding for interoperating with a service through messages (at a protocol level), which declares how to actually access it: the message formats, port numbers, and the way to exchange the input/output data. We will mainly use output and input of process model in service model. Figure 2 shows a segment of a service advertisement of selling books.

```

<?xml version="1.0" encoding="WINDOWS-1252"?>
.....
</profile:Profile>
<process:ProcessModel rdf:ID="BOOK_PRICE_PROCESS_MODEL">
<service:describes rdf:resource="#BOOK_PRICE_SERVICE"/>
<process:hasProcess rdf:resource="#BOOK_PRICE_PROCESS"/>
</process:ProcessModel>
<process:AtomicProcess rdf:ID="BOOK_PRICE_PROCESS">
<process:hasInput rdf:resource="#_TITLE"/>
<process:hasInput rdf:resource="#_PUBLISHER"/>
<process:hasOutput rdf:resource="#_AUTHOR"/>
<process:hasOutput rdf:resource="#_PRICE"/>
</process:AtomicProcess>
<grounding:WsdGrounding rdf:ID="BOOK_PRICE_GROUNDING">
.....
</grounding:WsdGrounding>
</rdf:RDF>

```

Fig. 2. A segment of book selling service

2.2 Chord-A kind of Structure P2P Networks

Chord [7] is a structured P2P network model which is based on consistent hashing. It has good expansibility, robustness, equality of node load and capacity of self-organization. In Chord network, the nodes are constructed to a circle in logic. Every node has an identifier which is usually named NID and every object has an identifier which is named the OID. NID usually can be provided by hashing information of the node itself and OID can be provided by hashing the name of object. The NID and the OID are mapped into the same space. An object is stored in the first node whose NID is greater or equal to its OID. In this paper a node represents a service registry and the object represents the registered service advertisement.

3 WordNet-Enhanced Semantic Web Service Discovery

As we have mentioned above, the centralized registries suffer from the problem of single point failure and performance bottleneck, while the decentralized unstructured registries suffer from the problem of high network consumption and lower precision. In order to overcome these problems, we select structured P2P networks as the foundation of Web services publication and discovery, and propose two improvements. Instead of the name of service advertisement, we select the output concept of service model in

OWL-S as the identifier of the service advertisement. When a Web service is to be published, the output concept of service is taken out and hashed, then the service is published to the corresponding registry. If the service has several output concepts, each concept should be hashed and the service will be published to several corresponding registries according to each concept. For example, if the service of selling books we given in section 2 is to be published, the concept AUTHOR and PRICE will be hashed separately. The service will be published to the registry according to AUTHOR and the registry according to PRICE.

The advantage is that the service advertisements are stored in several registries according to their output concepts and overcome the problem of single point failure. The services advertisements with same output concepts are stored in a registry and each registry only takes on service matchmaking with special output concepts that overcomes the problem of performance bottleneck. The registries communicate with each other through Chord. Compared with unstructured P2P networks using query flooding, the advantage of our method is that we can complete service routing and location in $O(\log_2 N)$ (N is the number of registries in the network) and finally complete the service discovery. This overcomes the problem of high network consumption and lower precision.

According to publication rule, the registries need some improvements. In registries the output concept of services is taken as Key and the list of service advertisements which can provide the output concept as Key value. The element of list is represented as $\langle \text{SID}, \{\text{input concepts}\} \rangle$, where SID represents the service identifier. For example, the Key value of output concept AUTHOR is $\{\langle \text{BOOK_PRICE_SERVICE}, \{\text{TITLE, PUBLISHER}\} \rangle, \dots\}$. If another service has the output concept AUTHOR, the service advertisement is added to the Key value of AUTHOR. With this improvement the registry is not the registry of service advertisement but the registry of output concepts.

Then we consider the situation of service discovery. At first according to the user's request we will get the output concept of service request and locate to the registry which stores the service advertisement. Then some relevant service matchmaking algorithms will be invoked to match the service advertisement to the service request. At present, a great many methods about service matchmaking are proposed in academia such as reference [8-10]. Reference [8] proposed a three level match method for semantic Web services, about text description, function description and service parameter. In this paper, we select this method for service matchmaking. It is notable that service matchmaking is not limited to any method. Every computation method can be used here. In this paper, the service matchmaking is just calculated between the input of service advertisement and service request because the output is matched at the location step. At last, the result returns to the user. Note that if the service request has several output concepts, our method will locate to several registries according to each concept and complete matchmaking in each registry. Then each registry returns the result. These results aren't the final results, but the results for each output concept of service request separately. The requester must calculate the intersection of these results and get the desired results. Another advantage of this method is that only input concepts are considered in matchmaking so that the complication of matchmaking is reduced greatly.

However, although our method overcomes some problems exist in the centralized registries and decentralized unstructured registries, the problem of lack of semantic information and only supporting exact searching isn't solved yet. For example the concept WRITER has the similar meaning of the concept AUTHOR. If we use hash function on these two concepts, obviously the value of hashing is different. If the output concept of service request is WRITER and the output concept of service advertisement is AUTHOR, the service matchmaking may fail although these two concepts have similar meaning because the service request will be located to the registry of WRITER not the registry of AUTHOR. In order to overcome this problem we propose the novel hashing technology based on synset of WordNet [11].

WordNet is a semantic dictionary-a lexical database, developed at Princeton by a group led by Miller. The version used in this study is WordNet 2.1 which has over 150,000 words organized in over 110,000 synonym sets.

WordNet partitions the lexicon into nouns, verbs, adjectives, and adverbs. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. All concepts in a synset have similar meaning. Thus, concepts in a synset are interchangeable in some syntax. Knowledge in a synset includes the definition of all these concepts. Synsets are interlinked by means of conceptual-semantic and lexical relations.

When a concept need to be hashed (in despite of service publication or service discovery), the synset of this concept is found out and then a special concept is chosen from the synset to generate the hash value. In order to generate unique hash value, we choose the first concept from all concepts sorted by letters in the synset. Since the result of sorting by letters is unique, it's sure that the hash value of the synset is unique.

4 Conclusions and Future Work

In this paper, we first introduced the technology of SOC and proposed two problems in Web services discovery: how to describe service accurately and delicately to support precisely matchmaking and how to store, index and exchange the information of Web services. Some methods have been proposed in order to solve the problems. However, unfortunately, there was some disadvantage in these methods more or less. In order to improve these methods we proposed a method for semantic Web services discovery using structured P2P technology. Furthermore, we proposed the hashing technology based on synsets of WordNet to overcome the problem of lack of semantic information and only supporting exact searching in structured P2P networks.

In practice, Web services discovery may not sufficient for the user's demand because of complicated application. How to compose such services in order to satisfy the high-level requirement especial in the distributed environment will be our future works.

References

1. Papazoglou, M.P., Georgakopoulos, D.: Service-oriented Computing. *Commun. ACM* 46, 25–28 (2003)
2. Chinnici, R., Moreau, J., Ryman, A., Weerawarana, S.: Web Services Description Language (WSDL) Version 2.0 Part 1: Core Language (W3C Recommendation) (June 26, 2007)

3. Mitra, N., Lafon, Y.: SOAP Version 1.2 Part 0: Primer. W3C Recommendation, 2nd edn. (April 27, 2007)
4. Universal Description Discovery and Integration (UDDI) Technical White Paper (last accessed June 8, 2007)
5. Castro, M., Druschel, P., Kermarrec, A., Rowstron, A.: One ring to rule them all: Service discovery and binding in structured peer-to-peer overlay networks. In: Proceedings of the SIGOPS European Workshop, France (2002)
6. Verma, K., Sivashanmugam, K., Sheth, A., Patil, A., Oundhakar, S.: METEOR-S WSDI: A Scalable P2P Infrastructure of Registries for Semantic Publication and Discovery of Web Services. *Journal of Information Technology and Management* 6, 17–39 (2005)
7. Stoica, I., Morris, R., Karger, D.: Chord: A scalable peer-to-peer lookup service for internet applications. In: Proceedings of the ACM SIGCOMM 2001 Conference, pp. 149–160. ACM, Santa Barbara (2001)
8. Zhou, N., Song, Z.L., Ai, W.H., Chen, L.: Research of Three Level Match about Semantic Web Services. *Journal of System Simulation* 21, 2081–2084 (2009) (in Chinese)
9. Calado, I., Barros, H., Bittencourt, I.: An approach for semantic web services automatic discovery and composition with similarity metrics. In: Proceedings of the 2009 ACM Symposium on Applied Computing, pp. 694–695. ACM, Honolulu (2009)
10. Junghans, M., Agarwal, S., Studer, R.: Towards practical semantic web service discovery. In: Aroyo, L., Antoniou, G., Hyvönen, E., ten Teije, A., Stuckenschmidt, H., Cabral, L., Tudorache, T. (eds.) *ESWC 2010. LNCS*, vol. 6089, pp. 15–29. Springer, Heidelberg (2010)
11. Miller, G.A.: WordNet: A Lexical Database for English. *Commun. ACM* 38, 39–41 (1995)

A New Scheduling Algorithm in Hadoop MapReduce

Zhiping Peng¹ and Yanchun Ma²

¹ Computer and Electronic Information College, Guangdong University of Petrochemical Technology, Maoming China

mmxypzhp@yahoo.com

² School of Computer Science, Jiangsu University of Science and Technology, Zhenjiang China

mayanchun_ah@163.com

Abstract. The core concept of cloud computing is the resource pool. Hadoop MapReduce is a software framework for easily writing applications which process vast amounts of data in-parallel on large clusters of commodity hardware in a reliable, fault-tolerant manner. A MapReduce job usually splits the input data-set into independent chunks which are processed by the map tasks in a completely parallel manner. We distribute the total slots according to P_i which is the percent of job's unfulfilled tasks in the total unfulfilled tasks. Since the P_i of the large job is bigger, the large job will be allocated more slots. We can clearly improve the response time of the large jobs. This new scheduling algorithm can improve the performance of the system, such as throughput, response time.

Keywords: Cloud computing, MapReduce, scheduling algorithm.

1 Introduction

As the product of many technology mixed evolution, cloud computing is rapid change with the promoting of many large companies. So far, cloud computing has not yet been defined unified. Cloud computing [1] is the next stage in evolution of the Internet. The cloud in cloud computing provides the means through which everything from computing power to computing infrastructure, applications, business processes to personal collaboration can be delivered to you as a service wherever and whenever you need.

The core concept of cloud computing is the resource pool. Task scheduling, one of the research fields of cloud computing, is still in the basic stage. We need to study it further to improve the efficiency of cloud computing systems.

Recently, a wide variety of companies and organizations use Hadoop for both research and production. Apache Hadoop is a framework for running applications on large cluster built of commodity hardware. The Hadoop framework transparently provides applications both reliability and data motion. Hadoop implements a computational paradigm named Map/Reduce, where the application is divided into many small fragments of work, each of which may be executed or reexecuted on any node in the cluster. In addition, it provides a distributed file system (HDFS) that stores data on the compute nodes, providing very high aggregate bandwidth across the cluster.

This paper is organized as follows. Section 2 describes Hadoop's scheduler. Section 3 introduces our new scheduling algorithm. Section 4 shows the result of our experiment.

2 Related Work

All schedulers in Hadoop provide access to a `TaskTrackerManager` --an interface to the `JobTracker` as well as a `Configuration` instance. Task assignment in Hadoop is reactive. `TaskTrackers` periodically send heartbeats to the `JobTracker` with their `TaskTrackerStatus`, which contains a list of running tasks, the number of slots on the node, and other information. The `JobTracker` then calls `assignTasks` on the scheduler to obtain tasks to launch. These are returned with the heartbeat response.

There are several scheduling algorithms for Hadoop which provides a way to share large clusters. We firstly introduce them as follows.

2.1 FIFO (First in First Out)

FIFO scheduling is the default method. All users' jobs are submitted to a single queue, they are sequentially executed according to their submission time. The implementation of FIFO is simple, and the whole cluster's scheduling overhead is less. However, the drawback of FIFO scheduling is the short job's response time is bigger while the large jobs exist. One user's large job should not monopolize the whole cluster, delaying the completion of everyone else's (small) jobs.

2.2 Capacity Scheduling

There are multiple job queues to maintain users' sub-mitted jobs. Each job queue contains a certain amount of tasks on the basis of the configuration. All job queues will be allocated reasonable resources according to configuration files. A new task is randomly submitted to one job queue.

Queues are allocated a fraction of the capacity of the grid in the sense that a certain capacity of resources will be at their disposal. All jobs submitted to a queue will have access to the capacity allocated to the queue. Free resources can be allocated to any queue beyond it's capacity. When there is demand for these resources from queues running below capacity at a future point in time, as tasks scheduled on these resources complete, they will be assigned to jobs on queues running below the capacity.

Within a queue, jobs with higher priority will have access to the queue's resources before jobs with lower priority. However, once a job is running, it will not be preempted for a higher priority job, though new tasks from the higher priority job will be preferentially scheduled.

2.3 Fair Scheduling

The Hadoop fair scheduler started as a simple means to share MapReduce clusters. Overtime, it has grown in functionality to support hierarchical scheduling, preemption, and multiple ways of organizing and weighing jobs.

Fair scheduling is a method of assigning resources to jobs such that all jobs get, on average, an equal share of resources over time. When there is a single job running,

that job uses the entire cluster. When other jobs are submitted, tasks slots that free up are assigned to the new jobs, so that each job gets roughly the same amount of CPU time. The fair scheduler organizes jobs into pools, and divides resources fairly between these pools. Within each pool, jobs can be scheduled using either fair sharing or First-in-first-out (FIFO) scheduling.

In addition to providing fair sharing, the fair scheduler allows assigning guaranteed minimum shares to pools, which is useful for ensuring that certain users, groups or production applications always get sufficient resources. When a pool contains jobs, it gets at least its minimum share, but when the pool does not need its full guaranteed share, the excess is split between other pools.

If a pool's minimum share is not met for some period of time, the scheduler optionally supports preemption of jobs in other pools. The pool will be allowed to kill tasks from other pools to make room to run. Preemption can be used to guarantee that "production" jobs are not starved while also allowing the Hadoop cluster to also be used for experiment-al and research jobs.

3 New Scheduling Algorithm

In this paper, we will introduce a new scheduling algorithm based on fair scheduling and compare it with fair scheduling in the efficiency.

We use resource pool to manage the submitted jobs. This new scheduling algorithm creates a resource pool for each user in case of default. It's a two level hierarchical scheduling algorithm. The first level: assigning slots between pools; the second level: assigning slots in its own pool.

3.1 Assigning Slots

A MapReduce job usually splits the input data-set into independent chunks which are processed by the map tasks in a completely parallel manner. The framework sorts the outputs of the maps, which are then input to the reduce tasks. There are M map tasks and R reduce tasks to be executed.^[2] The account of these tasks increases linearly with the size of the job. In the implementation process, we can acquire of the task's status, such as failed, succeed and skipped.

Assumed that here n jobs in the system, we count the number of the tasks whose status is not succeed as n_i .

$$sum = \sum_{i=1}^n n_i \quad (1)$$

$$P_i = \frac{n_i}{sum} \times 100\% \quad (2)$$

P is the percent of job's unfulfilled tasks in the total unfulfilled tasks.

We redistribute the total slots according to the P_i . Since the P_i of the large job is bigger, the large job will be allocated more slots. We can clearly reduce the response time of the large jobs. After the redistributions, each job's processing progress tends to the same. Clearly, the disadvantage is that small jobs' response time will be affected.

When a new job is submitted, we will update P_i . The new job's tasks have not been executed. And its P is 100%, so it will get some slots timely. And we can reduce its wait time. We also set a fixed time as a interval to recalculate the P_i .

In this scheduling algorithm, preemption is allowed. When we select the task to be skilled, the default choice is the recently started task in order to minimize the loss. Preemption does not lead to the failure of the preempted task, because Hadoop jobs admit the lost of the tasks.

3.2 Data Locality

When more small jobs are submitted, the possibility of data stored on this node is only 6%, the possibility on this rack is 61%^[3]. However, the most important demand is executing the tasks on the node which is close to the data storage node.

Improving the data locality can increase the system's throughput, because the network bandwidth in a large computer cluster is much smaller than the disk total I/O bandwidth between nodes in the cluster.

When some nodes release the free slots and the task in the head location of the queue does not meet the data locality, we allow the task to wait T_1 for the node meeting the node locality and wait T_2 for the node meeting the rack locality. Obviously, T_2 is bigger than T_1 . Depending on the size of the cluster and the job we set the wait time.

The algorithm is as follows:

```

var time_waited_for_local_Map=0,
skipped_at_last_heart_beat=false,
Locality_Level=0;
Switch (lastMapLocalityLevel) {
    Case 0:
        If (timeWaitedForLocalMap>=T2) {
            Locality_Level=2;
            Else if (timeWaitedForLocalMap>=T1) {
                Locality_Level=1;
            }
            Else {
                Locality_Level=0;
            }
        }
    Break;
    Case 1:
        If (timeWaitedForLocalMap>=T1) {
            Locality_Level=2;
        }
}

```

```

Else {
    Locality_Level=1;          }
Break;
Default:
    Locality_Level=2;
}

```

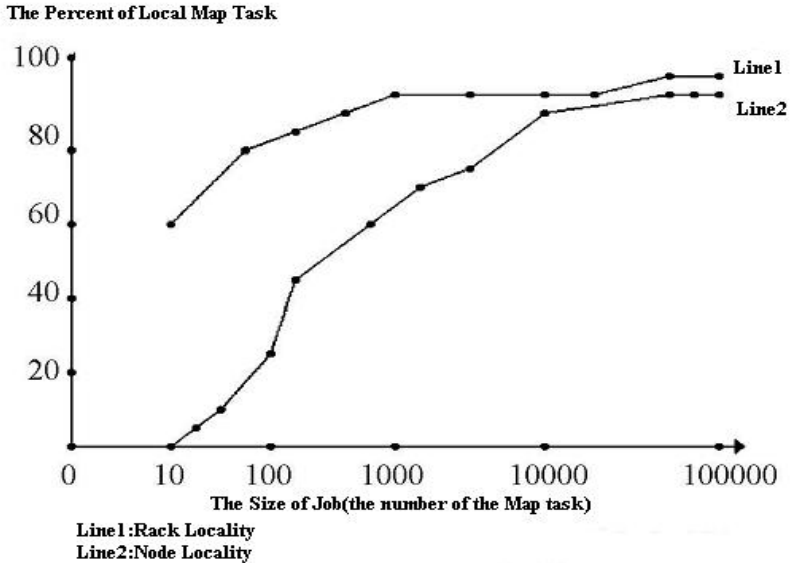


Fig. 1. The Relation between locality and job's size

If the task's locality_level is 0, it is allowed to be executed on the node which has the data copy. If its locality_level is 1 and it can be executed on the node in the rack which has the data copy. If its locality_level becomes 2 and it can be executed on any node.

4 Experiments

We performed our experiments on a private cluster which consists of 60 nodes. And 60 nodes are distributed into 3 racks. Hadoop's version is Hadoop-0.21.0. One node is set as the Job Tracker and NameNode, other nodes is Task Tracker and DataNode.

We implemented FIFO scheduling, fair scheduling and our new scheduling and compared their average response time by different sizes of jobs.

We consider a small set of applications, but we can run each application with a variety of inputs. A particular application reading from a particular set of inputs is called an application instance. A given application instance always uses the same

input partitions across all experiments. The applications in our experiments are as follows:

- **Sort:** This application sorts a set of 100 byte records that have been distributed into n partitions. The application starts by reading all the input data, sampling it to compute balanced ranges, and then range partitioning it into S partitions. Sort redistributes all of its input data between computers during this repartitioning and is therefore inherently network-intensive. Once the data has been repartitioned each computer sorts its range and writes its output back to the cluster.
- **WordCount:** This application reads text files and counts how often words occur. The input is text files and the output is text files, each line of which contains a word and the count of how often it occurred, separated by a tab. There are ten instances with inputs divided into 2, 4, 5, 6, 8, 10, 15, 20, 25, and 100 partitions respectively, and each partition is about 50MB in size. WordCount performs extensive data reduction on its inputs, so if it is scheduled optimally it transfers very little data over the network.
- **Grep:** This application extracts matching strings from text files and counts how many times they occurred. The program runs two map/reduce jobs in sequence. The first job counts how many times a matching string occurred and the second job sorts matching strings by their frequency and stores the output in a single output file.

By analyzing the results we found our new scheduling can clearly decrease the response time in comparison with FIFO scheduling. And in most cases, our new scheduling can achieve the same effect with fair scheduling. When many large jobs are submitted, the new scheduling can get a better average response time than others.

Acknowledgment. We would like to thank our colleagues for their many helpful comments and discussions. We would also like to the ICIS committee for their detailed reviews.

References

1. Hurwitz, J., Bloor, R., Kaufman, M., Halper, F.: *Cloud Computing For Dummies*, Hoboken (2009)
2. Apache.: Welcome to Apache Hadoop, <http://hadoop.apache.org/>
3. Wang, K., Wu, Q., Yang, S.: Design and Implementation of Job Scheduling Algorithm for Multi-User Map-Reduce Clusters. *J. Computer and Modernization* 8, 23–28 (2010)
4. Isard, M., Prabhakaran, V., Currey, J., Wieder, U., Talwar, K., Goldberg, A.: Quincy: Fair Scheduling for Distributed Computing Clusters. In: *ACM SIGOPS 22nd Symposium on Operating Systems Principles*, pp. 261–276. ACM Press, Montana (2009)
5. Dean, J., Ghemawat, S.: Mapreduce: simplified data processing on large clusters. *Communications of the ACM*, 107–113 (2008)
6. Feitelson, D.G., Rudolph, L., Schwiegelshohn, U.: Parallel job scheduling—a status report. *Computer Science* 3277, 1–16 (2005)
7. Ucar, B., Aykanat, C., Kaya, K.: Task assignment in heterogeneous computing systems. *Parallel and Distributed Computing* 66, 32–46 (2006)
8. Graves, S.C., Redfield, C.H.: Equipment Selection and Task Assignment for Multiproduct Assembly System Design. *Flexible Manufacturing Systems* 1, 31–50 (1988)

9. Harchol-Balter, M.: Task assignment with unknown duration. *JACM* 49, 266–280 (2002)
10. Kettimuthu, R., Subramani, V., Srinivasan, S., Gopalsamy, T., Panda, D.K., Sadayappan, P.: Selective preemption strategies for parallel job scheduling. *High Performance Computing and Networking* 3, 122–152 (2005)
11. Aida, K.: Effect of job size characteristics on job scheduling performance. *Computer Science* 1911, 1–17 (2000)
12. Nicolae, B., Moise, D., Antoniu, G., Bouge, L.: BlobSeer: Bringing high throughput under heavy concurrency to Hadoop Map-Reduce applications. In: *Parallel & Distributed Processing (IPDPS)*, pp. 1–11. IEEE Press, Atlanta (2010)
13. Bhandarkar, M.: MapReduce programming with apache Hadoop. In: *Parallel & Distributed Processing (IPDPS)*, p. 1. IEEE Press, Atlanta (2010)
14. Yeung, J.H.C., Tsang, C.C., Tsoi, K.H., Kwan, B.S.H., Cheung, C.C.C., Chan, A.P.C.: Map reduce as a Programming Model for Custom Computing Machines. In: *Field-Programmable Custom Computing Machines*, pp. 149–159. IEEE Press, Palo Alto (2008)

The Kinetics Simulation of Robot Based on AutoCAD

Yu Ding

Chongqing City Management College
Chongqing, China
40129784@qq.com

Abstract. The kinetics simulation is the key point of robot design. A robot kinetics simulation system based on AutoCAD is developed in this paper. By the Visual LISP which is in the Auto CAD software, the various robot models are obtained. The types and the size of robots can be arbitrarily chosen by user according to the practical requirement. The motion simulation of robots can be achieved in this system, and the parameters of kinetics also obtained with solving the kinetics equation. Based on this system, a typical robot is analyzed and the satisfied kinetics solutions are obtained. The results show this system is a well off-line programming simulation system.

Keywords: Kinetics, Simulation, CAD, LISP.

1 Introduction

The geometrical model of robot is constructed by using the knowledge of computer graphics and driven by program. If the parts don't interfere, the program will be transferred to the robot. This method is off-line programming. It has many advantages in improving the efficiency and cutting the cost. The key points in off-line programming are 3D parametric modeling and driving. Visual LISP is the list process language that is inlay in AutoCAD. This language has strong function in parametric modeling [1]. So we establish a robot simulation system with Visual LISP. The main functions of system include: 3D parametric modeling; robot kinetics simulation; checking of interference; calculating the kinetics trail of robot.

2 The Interface of System

The Fig. 1 shows the interface of system. This custom-built menu in AutoCAD is realized by modifying the "acad.mnu" file [2]. The file will be loaded when AutoCAD is activated. The menu includes: robots modeling, kinetics simulation, edit tools, help. The robots menu have sub-menu that are used to model the parts.

3 Parametric Modeling

3.1 The Program of Modeling

We think that all robots consist of frame, pillar, arm, wrist and robot manipulator. For those parts, we use Visual LISP program to construct them. Because the parts are

simpler, the modeling program is simpler. We only transfer some base commands in AutoCAD, e.g. “box”, ”cylinder”, “extrude”, “section”, etc [3]. Based on those AutoCAD commands and some Visual LISP commands, parts of different size and shape will be obtained. And then, by some characteristic point and “osnap” command, we can assemble those parts according to different robots structure. Some results are shown in Fig. 2.



Fig. 1. The interface of system by custom built

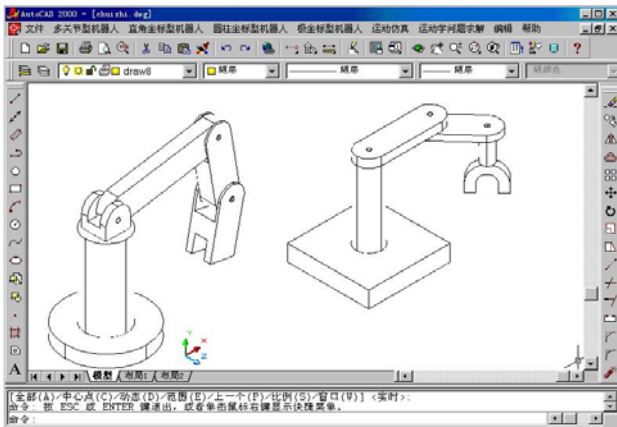


Fig. 2. The parts model of some robots

3.2 Parameters Inputting

The Visual LISP program of modeling can construct the drawing in automation. The parameters inputting actualized by the dialogue control language which also inlaid in AutoCAD. Using the language, we build some dialogue interface. So the parameters inputting become very effective, convenient and accurate. In this system, all parameter are inputted in dialogue box. The Fig. 3 shows an example.

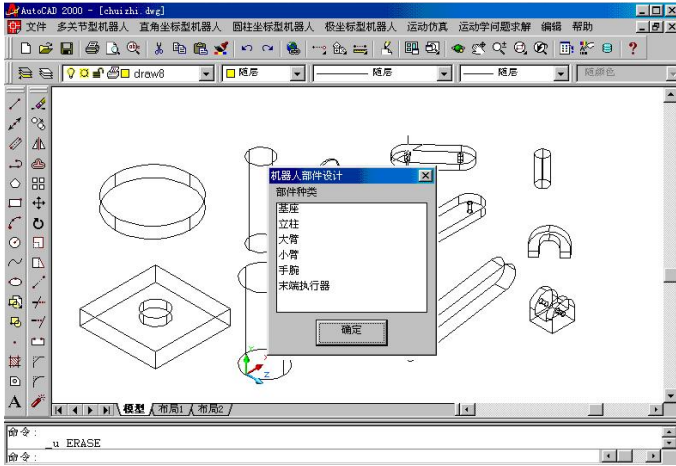


Fig. 3. The dialogue box of robot modeling

4 The Simulation of Robot

4.1 Driven

Because the AutoCAD and Visual LISP have no motion function, how to actualize the kinetics simulation of robot is the key point.

At first, we establish some dialogue for input parameters that obtained by solutions for inverse equations of robot. The parameters may be inputted automatically, also can be inputted manually. After the parameters are inputted, some selection sets will be built by Visual LISP program. And then, the selection sets will be transferred into some cycle

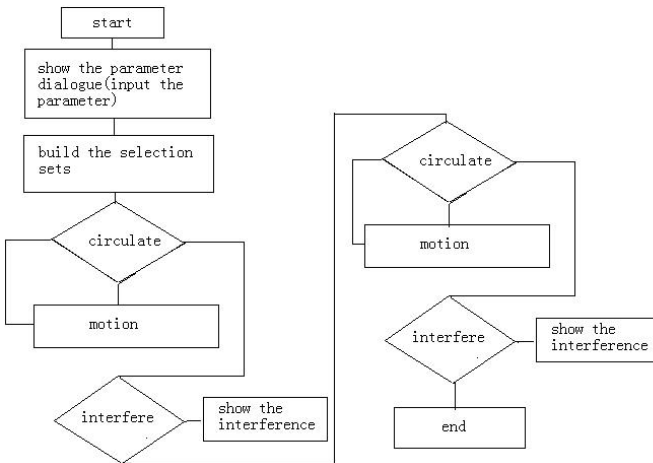


Fig. 4. The diagram of motion program

functions, e.g. “while”, “repeat”, “foreach”. The substances of cycle functions are “rotate” or “move” commands in AutoCAD. Depending on those commands, all parts of robots will be driven. The process of driven program is shown in Fig. 4.

4.2 The Examination of Interference

The kinetics simulation of robot has another important aspect ---- the examination of interference. We examine the interference of all parts depending on “interfere” command that will be written in program [4]. When the parts move or rotate one step, the “interfere” command will be performed once. If some parts are interfering, the results will be shown. To make the examining work convenient and rapid, we construct the parts in different colors. So the color of interference will be white which can be select rapidly by using “ssget” command in Visual LISP. And then the selection will be shown in corresponding position. This examination is very important to the simulation of robot. The process of program is also shown in Fig. 4.

5 An Example of the Simulation of Robot

We use a robot as an example [5], establish the mathematical model that means the relative position of all parts. The structure and D-H coordinate of robot is shown in Fig. 5. h_i is the length of parts, β_i is the angle of rotation.

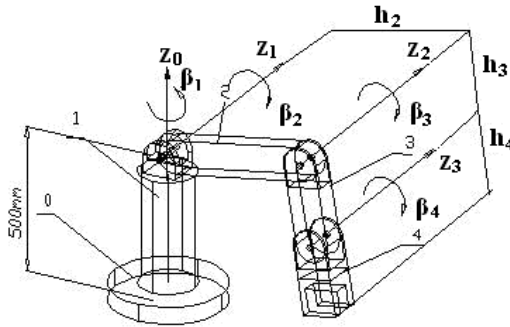


Fig. 5. The structure and D-H coordinate of robot

The relationships between the center of robot manipulator and center of frame are obtained by matrix X [6],

$$X=M04 \times E \tag{1}$$

where $M04$ is obtained by:

$$M04=[M01 M12 M23 M34] \tag{2}$$

where

$$M_{01} = \begin{bmatrix} \cos \beta_1 & 0 & -\sin \beta_1 & 0 \\ \sin \beta_1 & 0 & \cos \beta_1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$M_{12} = \begin{bmatrix} \cos \beta_2 & -\sin \beta_2 & 0 & h_2 \cos \beta_2 \\ \sin \beta_2 & \cos \beta_2 & 0 & h_2 \sin \beta_2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$M_{23} = \begin{bmatrix} \cos \beta_3 & 0 & -\sin \beta_3 & h_3 \cos \beta_3 \\ \sin \beta_3 & 0 & \cos \beta_3 & h_3 \sin \beta_3 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$M_{34} = \begin{bmatrix} \cos \beta_4 & 0 & \sin \beta_4 & 0 \\ \sin \beta_4 & 0 & -\cos \beta_4 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$E = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

If we know the parameters of structure, the motion parameters will be solved and the X will be obtained. In this example, the $h_1=500$ mm, $h_2=500$ mm, $h_3=300$ mm, $h_4=300$ mm, $\beta_1 = 2\pi$, $\beta_2 = \pi$, $\beta_3 = \pi$, $\beta_4 = \pi$. So, X is

$$X = \begin{bmatrix} 2.462 & 0.236 & 0.594 & 548.5628 \\ 0.754 & 0.265 & 0.463 & -124.438 \\ -0.864 & 0.264 & 0.769 & 120.579 \\ 0.000 & 0.000 & 0.000 & 1.000 \end{bmatrix} \quad (3)$$

Equation (3) shows the position of robot manipulator.

6 Conclusions

The system will be an off-line programming system, when the drive program is loaded to the robot and transferred into the robot programming commands. It is

important to improving the efficiency, saving the debugging times and avoiding interfering. In present, a few of excellent simulation software are working. Visual LISP and AutoCAD are not the best choices. The research of this paper is only a branch of them. This system is meaningful to application of Visual LISP language.

References

1. John, W.: AutoCAD2000 3D modeling. China Machine Press, Beijing (2000)
2. Chen, D.J., Fu, S.M.: The application of Visual LISP. Sichuan University Press, Chengdu (1995)
3. Sun, J.H., Mi, J.: Programming and application of VisualLisp. Science Press, Beijing (1999)
4. Wang, Q.Z., Xu, X.H.: A new inferential method and efficient solutions for inverse kinematics equations of Puma robot manipulator. Robot 2, 81–87 (1998)
5. Liu, S.M., Chen, Y., Zhang, W.J.: Modular robots and computer aided design. Robot 3, 16–20 (1999)
6. Fu, K.S., Gonzalez, R.C., Lee, C.S.: Robotics. Science Press, Beijing (1989)

An Improved Particle Swarm Optimization Algorithm*

Dazhi Pan^{1,2} and Zhibin Liu²

¹ College of Mathematic & Information China West Normal University,
Nanchong 637002, P.R. China

² College of Mathematic & Physics Southwest Petroleum University,
Chengdu 610500, P.R. China
pdzzj@126.com

Abstract. An improved particle swarm optimization (IPSO) is proposed in this paper. In the new algorithm, the Optimal and sub-optimal locations, which each particle encountered and the swarm meted, are kept. Based on these locations, the IPSO produces four velocities for each particle and obtains the particle's iterative position. The IPSO enlarges the search space and enhances global search ability. From the results obtained through benchmark problems, it is clearly seen that the proposed IPSO method is competitive to the PSO.

Keywords: particle swarm optimization, improved particle swarm optimization, sub-optimal location, global optimization.

1 Introduction

Particle swarm optimization technique is considered as one of the modern heuristic algorithms for optimization introduced by James Kennedy and Eberhart[2,5]. It is based on the social behavior metaphor [2] and a population-based optimization technique, as an alternative tool to genetic algorithms and gained lots of attention in various optimal control system applications. PSO is a stochastic search technique with reduced memory requirement, computationally effective and easier to implement compared to other evolutionary algorithms. One key advantage is that PSO has memory, i.e., every particle remembers its best solution (local best) as well as the group's best solution (global best).As such PSO is well suited to tackling dynamical problems, another advantage of PSO is that its initial population is maintained fixed throughout the execution of the algorithm, and so, there is no need for applying operators to the population, a process which is both time- and memory-storage-consuming. But PSO exist some defects such as premature convergence, so this paper proposes a improved PSO algorithm.

In the new PSO, each particle saves the local Optimal and sub-optimal locations, while the group remembers the global Optimal and sub-optimal locations. For each

* Supported by the National Natural Science Foundation of China (Number:50874094).

particle, four saved positions are used to produce the next iterative velocity. By this way, the search scope can be expanded and the speed of convergence can be improved.

2 Review of PSO Algorithm

In 1995, Dr. Kennedy and Dr. Eberhart[1] introduced PSO, in which a swarm consists of a set of particles moving around the search space, each representing a potential solution. Each particle has a position vector ($x_i(t)$), a velocity vector ($v_i(t)$), the position at which the best fitness ($pbset_i$) encountered by the particle, and the index of the best particle ($gbest$) in the swarm.

In each generation, the velocity of each particle is updated according to their local best position and the global best using (1).

$$v_i(t+1) = v_i(t) + c_1 r_1(t)(pbset_i - x_i(t)) + c_2 r_2(t)(gbest - x_i(t)) \quad (1)$$

Where c_1 and c_2 are constants and are called acceleration co-efficient, namely cognitive and social parameter, respectively. $r_1(t)$ and $r_2(t)$ are uniformly-distributed random number in the range [0,1]. The position of each particle is updated every generation. This is done by adding the velocity vector to the position vector, as give in (2).

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (2)$$

However, in the first version of PSO, there was no actual control over the previous velocity of the particles. In the late version of PSO, the shortcoming was addressed by incorporating a parameter, called inertia weight introduced by Shi and Ebherhart[6]. The inertia weight is employed to control the impact of the previous history of velocities on the current one. See the following equation.

$$v_i(t+1) = wv_i(t) + c_1 r_1(t)(pbset_i - x_i(t)) + c_2 r_2(t)(gbest - x_i(t)) \quad (3)$$

Where w is called inertia weight, Notes here that the big w can help in discovering new solution spaces at the cost of a slower convergence rate, and the small w can contribute to find a better solution in current space at expense of missing diversity.

The PSO algorithm is further improved through using a time-decreasing inertia weight, which leads to a reduction in the number of iterations [4, 6]. The performance of this modified algorithm depends on the method of tuning the inertia weight. In this work, a linearly-decreasing time-dependent inertia weight proposed in [6] has been implemented, according to the following update equation.

$$w_t = (w_2 - w_1) \frac{m-t}{m} + w_1 \quad (4)$$

Where w_1 and w_2 are the lower and higher inertia weight values, m is the maximum number of iteration and t is the current number of iteration.

Recent works in [1] and [3] indicate that the use of a “constriction factor” may be necessary to insure convergence of the PSO. A simplified method of incorporating a constriction factor is represented in [3].

$$v_i(t + 1) = K[v_i(t) + \varphi_1 r_1(t)(pbest_i - x_i(t)) + \varphi_2 r_2(t)(gbest - x_i(t))] \tag{5}$$

Where K is a function of φ_1 and φ_2 as illustrated by the following equation.

$$K = \frac{2}{|2 - \varphi - \sqrt{\varphi^2 - 4\varphi}|}, \varphi = \varphi_1 + \varphi_2, \varphi > 4 \tag{6}$$

3 An Improved Particle Swarm Optimization (IPSO) Algorithm

Inspired by the swarm intelligence of particle swarm, a new variation of PSO is proposed in this paper. The new approach, called IPSO, expands the scope of excellent positions which are saved by each particle and the swarm, in order to improve the performance of PSO.

3.1 Generating the Particle’s Iterative Velocity

In order to take full advantage of the excellent positions encountered by each particle and the swarm, in the IPSO, the local best position of each particle and the group’s global best position are not only saved, but also the local second best position($spbest_i$) and the global second best position($sgbest$) are remembered, so there are four iterative velocities which are constructed by $pbest_i$, $spbest_i$, $gbest$ and $sgbest$. They are illustrated by the following equations.

$$v_i^1(t + 1) = wv_i(t) + c_1 r_{11}(t)(pbest_i - x_i(t)) + c_2 r_{21}(t)(gbest - x_i(t)) \tag{7}$$

$$v_i^2(t + 1) = wv_i(t) + c_1 r_{12}(t)(spbest_i - x_i(t)) + c_2 r_{22}(t)(gbest - x_i(t)) \tag{8}$$

$$v_i^3(t + 1) = wv_i(t) + c_1 r_{13}(t)(pbest_i - x_i(t)) + c_2 r_{23}(t)(sgbest - x_i(t)) \tag{9}$$

$$v_i^4(t + 1) = wv_i(t) + c_1 r_{14}(t)(spbest_i - x_i(t)) + c_2 r_{24}(t)(sgbest - x_i(t)) \tag{10}$$

The i th particle’s iterative velocity is given by following equation.

$$v_i(t + 1) = \{v_i^j(t + 1) | f(x_i(t) + v_i^j(t + 1)) = \max_{1 \leq k \leq 4} f(x_i(t) + v_i^k(t + 1))\} \tag{11}$$

Where $f(x)$ is the fitness value at x position.

3.2 Optimal and Sub-optimal Location Updating

After the $t+1$ th iterative position of each particle has been attained, the particle’s $pbest_i$ and $spbest_i$ should be renewed, while the group’s $gbest$ and $sgbest$ are updated. Their renewing equations are in following.

$$spbest_i = \begin{cases} pbset_i, & \text{if } f(x_i(t+1)) > f(pbset_i) \\ x_i(t+1), & \text{if } f(pbset_i) \geq f(x_i(t+1)) > f(spbest_i) \\ spbest_i, & \text{if } f(x_i(t+1)) \leq f(spbest_i) \end{cases} \quad (12)$$

$$pbset_i = \begin{cases} pbset_i, & \text{if } f(x_i(t+1)) \leq f(pbset_i) \\ x_i(t+1), & \text{if } f(x_i(t+1)) > f(pbset_i) \end{cases} \quad (13)$$

$$sgbest = \begin{cases} gbest, & \text{if } f(pbset_i) > f(gbest) \\ pbset_i, & \text{if } f(gbest) \geq f(pbset_i) > f(sgbest), i = 1, \dots, N. \\ gbest, & \text{if } f(pbset_i) \geq f(sgbest) \end{cases} \quad (14)$$

$$gbest = \begin{cases} gbest, & \text{if } f(pbset_i) \leq f(gbest) \\ pbset_i, & \text{if } f(pbset_i) > f(gbest) \end{cases}, i = 1, \dots, N. \quad (15)$$

Where N is the number of particles in the particle swarm.

3.3 Algorithm of the Improved PSO

The improved PSO algorithm can be described as following.

Step1: initialization: setting the max iterative steps and the number of the particles, producing randomly the position and velocity of each particle in the particle swarm;

Step2: Evaluate the fitness value of each particle;

Step3: According to (12) and (13), renew the local optimal and sub-optimal position of each particle;

Step4: According to (14) and (15), renew the global optimal and sub-optimal position in the particle swarm;

Step5: Change the velocity of each particle according to (7-11);

Step6: Move each particle to the new position according to (2) and return to Step 2;

Step7: Repeat Step2–Step6 until a stopping criterion is satisfied.

4 Experimental Results and Analysis

In order to validate the performance of the improved particle swarm optimization algorithm, in this section, three nonlinear benchmark functions that are commonly used in literature [7-8] are performed. The functions, the admissible range of the variable and the optimum are summarized in following.

(1) Rastrigrin function

$$f_1(x) = \sum_{i=1}^n (x_i^2 - 10 \cos(2\pi x_i) + 10)$$

The admissible range of the Rastrigrin’s variable is $[-5.12, 5.12]^n$, and the optimum is 0.

(2)Rosenbrock function

$$f_2(x) = \sum_{i=1}^{n-1} (100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2)$$

The admissible range of the Rsenbrock’s variable is $[-30, 30]^n$, and the optimum is 0.

(3)Griewark function

$$f_3(x) = \frac{1}{400} \sum_{i=1}^n x_i^2 - \prod_{i=1}^n \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$$

The admissible range of the Griewark’s variable is $[-600,600]^n$, and the optimum is 0.

To evaluate the performance of the proposed IPSO, the basic PSO is used for comparisons. For all these methods, the parameter w used is recommended from Shi and Eberhart [6] with a linearly decreasing (see (4)), which changes from 0.9 to 0.4. The acceleration constants c_1 and c_2 are both 2.0. V_{max} is to equal X_{max} .

Table 1. Mean function values of Rastrigrin

Population Size	Dimension	Generation	PSO	IPSO
20	10	1000	5.2062	3.1248
	20	1500	22.7724	15.4329
	30	2000	49.2942	34.7920
40	10	1000	3.5697	2.4537
	20	1500	17.2975	13.9325
	30	2000	38.9142	25.8346
80	10	1000	2.3835	1.6098
	20	1500	12.9020	7.1053
	30	2000	30.0375	12.9857
160	10	1000	1.4418	0.8101
	20	1500	10.0438	4.3792
	30	2000	24.5105	12.4387

As in [7], for each function, three different dimension sizes are tested. They are dimension sizes: 10, 20 and 30 and the corresponding maximum number of generations are set as 1000, 1500 and 2000. In addition, different population sizes are used for each function with different dimensions. They are population sizes of 20, 40, 80, and 160. In this paper, this Experimental program is adopted for comparison of two algorithms’ performance. We execute the there algorithm in 30 independent runs. The mean fitness values of the best particle found for the 30 runs for the three functions are listed in Tables 1–3.

From table1-table3, it can be seen that the performance of IPSO is better than PSO's, so the improvement of PSO is effective and feasible.

Table 2. Mean function values of Rosenbrock

Population Size	Dimension	Generation	PSO	IPSO
20	10	1000	42.6162	33.5627
	20	1500	87.2870	67.9324
	30	2000	132.5973	90.7649
40	10	1000	24.3512	20.7482
	20	1500	47.7243	36.8739
	30	2000	66.6341	51.4326
80	10	1000	15.3883	13.4789
	20	1500	40.6403	32.5904
	30	2000	63.4453	57.4736
160	10	1000	11.6283	8.0826
	20	1500	28.9142	19.4732
	30	2000	56.6689	40.6826

Table 3. Mean function values of Griewark

Population Size	Dimension	Generation	PSO	IPSO
20	10	1000	0.0920	0.0721
	20	1500	0.0317	0.0257
	30	2000	0.0482	0.0298
40	10	1000	0.0762	0.0592
	20	1500	0.0227	0.0164
	30	2000	0.0153	0.0132
80	10	1000	0.0658	0.0470
	20	1500	0.0222	0.0184
	30	2000	0.0121	0.0097
160	10	1000	0.0577	0.0383
	20	1500	0.0215	0.0195
	30	2000	0.0121	0.0099

5 Conclusions

This paper presented an improved particle swarm optimization algorithm. The proposed IPSO algorithm expands the saving scope of excellent positions of the swarm and improves the performance of PSO. In IPSO, we let each particle and the swarms to remember their Optimal and sub-optimal locations, which enlarges the search space and enhances global search ability. Form the Experimental results, IPSO is efficient.

References

- [1] Clerc, M., Kennedy, J.: The particle swarm: Explosion, stability, and convergence in a multi-dimensional complex space. *IEEE Trans. Evol. Comput.* 6, 58–73 (2002)
- [2] Eberhart, R.C., Kennedy, J.: A new optimizer using particle swarm theory. In: *Proceedings of the Sixth International Symposium on Micro machine and Human Science, Nagoya, Japan*, pp. 39–43 (1995)
- [3] Eberhart, R.C., Shi, Y.: Comparing inertia weights and constriction actors in particle swarm optimization. In: *Proc. 2000 Congr. Evol. Comput.*, pp. 84–88 (2000)
- [4] Hu, X., Shi, Y., Eberhart, R.: Recent advances in particle swarm. In: *Proc. IEEE Congr. Evol. Comput.*, vol. 1, pp. 90–97 (June 2004)
- [5] Kennedy, J., Eberhart, R.C.: Particle swarm optimization. In: *Proceedings of IEEE International Conference on Neural Networks, Piscataway, NJ*, pp. 1942–1948 (1995)
- [6] Shi, Y., Eberhart, R.C.: A modifid particle swarm optimizer. In: *Proceedings of the IEEE Congr. Evol. Comput, Piscataway, NJ*, pp. 69–73 (1998)
- [7] Shi, Y., Eberhart, R.C.: Empirical study of particle swarm optimization. In: *Proceedings of the IEEE Congr. Evol. Comput., Piscataway, NJ*, pp. 1945–1950 (1999)
- [8] Trelea, I.C.: The particle swarm optimization algorithm: convergence analysis and parameter selection. *Information Processing Letters* 85, 317–325 (2003)

An Improved Optimal Matching for Video Retrieval

Hu Shuangyan, Li Junshan, and Feng Fujun

The Xi'an High-Tech of Institute, 710025, China

Abstract. Traditional methods and techniques of retrieval are inefficient on vary large number of video data. But, the content-based retrieval shows a right direction for information retrieval. Analysis of the features of maximum matching and optimal matching, the former reflects only from the perspective of the visual similarity of the two shots, and the latter objectively point of view fully reflected the contents of the similar characteristics of the shots on the basis of maximum matching from, but the most optimal matching algorithm ignores the time sequence of video content. This paper presents an improved optimal matching algorithm, and the experimental results showed the algorithm in the recall and precision rates are both obtain more satisfactory results.

Keywords: video retrieval, maximum matching, optimal matching.

1 Introduction

With the rapid development of image analysis and video processing technology, digital video applications has grown from the original video recording and video surveillance, intelligent video applications began to develop. The face of massive videos data storage management and retrieval applications, especially content-based analysis and retrieval applications, content-based video retrieval technology (CBVR) came into being [1].

Content-based image retrieval technology is a fuzzy query technology, by extracting visual features of the image, for example, color, texture, shape, spatial relations, etc., semantic features, like feature semantics, object semantics and abstract semantics[2-3], to find the closest image in the feature space[1]. In order to achieve large-scale video database for video data retrieval techniques and methods, this technology automatically extract and describe the characteristics and content of the videos without human intervention.

2 Video Retrieval

Shot boundary detection is the general method of video retrieval, and shots as the basic structural unit of the video series and retrieval unit. And then, extracts the key-frames within each shot to represent the content of the shots. Finally, using the colors, textures and shape features of key-frames to retrieval videos based on shots. Literature [4] gives a shot retrieval method which based on the best match of bipartite

graph. The similarity of the two shots is modeled as a weighted measure of the bipartite graph, frame in the shots as a bipartite graph node, the similar values of any frames between the two shots as edge weights, in one to one premise, using the optimal matching algorithm obtained the maximum weight bipartite graph as the similarity between the two shots. In order to improve retrieval speed, literature [4] proposed two improved algorithms, sub-shot constructed weighted bipartite graph and spaced sampling constructed weighted bipartite graph. However, the algorithm of literature, have some limitations, [4] is based on the movement of the camera, and do not fully take into account changes between content of scene and camera motion. For example, literature [4] will be failed when the camera is stationary and the content of the scene is changed. In addition, spaced sampling used to construct the weighted bipartite graph so that the two shots have the same number of key-frames, produces data redundancy and increases the complexity of matching calculating, and the Contribution to the final similarity is very limited.

According to the matching theory of graph theory, in order to achieve fast and efficient shot retrieval, many shots which have the smaller similarity should be filtered out, and we can get it by rough matching with the maximum matching and using optimal matching algorithm to get the final similarity of shots. In this paper, it gives the definition of the similarity and matching of shots. The maximum matching and optimal matching is combined to achieve content-based video shot retrieval, and partial improved in optimal matching. The series features are introduced to the retrieval to improve the accuracy of the algorithm.

2.1 Maximum Matching and Optimal Matching

After the key-frame sequence obtained, the video shot can be represented by them. Set the key-frame sequence of the query shot is $Q = \{q_1, q_2, \dots, q_n\}$, the shot in video-base is expressed as $D = \{d_1, d_2, \dots, d_m\}$, where q_i and d_i are key-frames. Typically, the key-frame numbers of different shot is not the same.

The similar condition of two shot is to find the pairs of key-frames, the correspondence between key-frames are one to one, and then get the largest similarity of the shots. To solve this problem, the ideas of two graph theory, maximum matching and optimal matching, can be used to shot-based video retrieval. The maximum matching and optimal matching are used in the video clip retrieval by literature [5] and had achieved good results. But it applied the maximum matching and optimal matching algorithm only from the perspective of graph theory, the video shots is simply equivalent to a set of video frames, not considered the characteristics of the video itself.

The maximum matching and optimal matching are different in properties, the maximum matching reports the similarity of the two shots only in vision, the optimal matching stand on the maximum matching and objectively and comprehensively reports the similar characteristics of shots content, and the computational complexity of maximum matching less than the optimal matching.

Therefore, using the maximum matching at first, can quickly get the set of similar shots, and then use the optimal matching to calculate the similarity of two video shots.

Maximum matching and optimal matching are two classic problems of graph theory. Suppose there are $n (x_1, x_2, \dots, x_n)$ workers to complete $m (y_1, y_2, \dots, y_m)$ tasks, and each worker does at most one of the tasks involved, and each task can only be completed by one worker. On this issue, maximum matching problems how to make as many as people have work to do and optimal matching problems how to make reasonable arrangements for the highest overall efficiency. Fig.1 and Fig.2 show the theory of maximum matching and optimal matching.

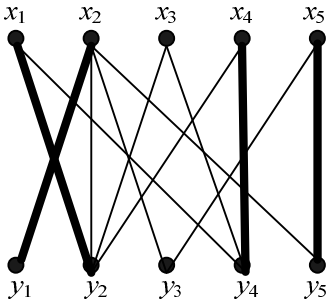


Fig. 1. Maximum matching

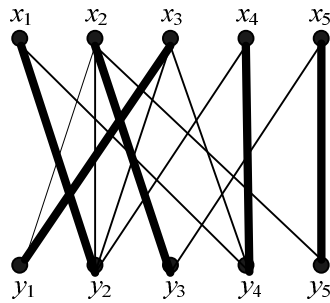


Fig. 2. Optimal matching

The optimal matching of bipartite graph is used in shot retrieval, and the video shot retrieval can be described as: suppose the key-frame number of shot X is $n (x_1, x_2, \dots, x_n)$, shot Y in video base have $m (y_1, y_2, \dots, y_m)$ key-frames, the edge sets $e_{x_i, y_j} = (x_i, y_j)$ express the similarity between x_i and y_j , and e_{x_i, y_j} is the similarity value.

Hungarian algorithm [6] is a classical maximum matching algorithm in graph theory. Maximum matching algorithm only solves the problem that make as much as possible key-frames matching. However, the video shot similarity retrieval is not only a simple pairing, also need to consider the similarity of key-frames. Kuhn-Munkres algorithm [6] is a classical algorithm to achieve optimal matching of bipartite graph. The basic idea is following. First, weighted complete bipartite graph G a feasible node label l , and then constructing the equivalent subgraph G_l according l , then find an exact matching of G_l with the exact match algorithm (do not need to consider the weight of each edge in this step), if there exist a an exact matching M^* , and M^* is the optimal matching of G , then the algorithm is end. Otherwise, modify the node label l , reconstruct the equivalent subgraph until it contains an exact matching.

Kuhn-Munkres algorithm neglects the time sequential of video. In a shot, video frames have a specific time sequence. So the optimal matching should also fully consider the time sequence. That is to say, back-tracking is an error in matching. For example, the two points x_3 and y_1 in figure 2 is one back-tracking. This matching is

right in graph theory, but the matching result maybe wrong when points have a special physical mean. That the output of Kuhn-Munkres algorithm is further processing can eliminate back-tracking.

Generally, back-tracking always considered in the reverse order and the idea is a one-sided. Therefore, a suit definition of back-tracking is given.

Definition 1: Given an optimal matching $G = (X, \Delta, Y)$, the edges in Δ can be divided into tow orderly sets Δ_1 and Δ_2 , where $\Delta_1 = \{e_{x_i, y_j} \mid x_i \leq y_j, x_i \in X, y_j \in Y\}$ and $\Delta_2 = \{e_{x_i, y_j} \mid x_i > y_j, x_i \in X, y_j \in Y\}$.

$$\begin{cases} \delta_1 = \sum e_{x_i, y_j} & e_{x_i, y_j} \in \Delta_1 \\ \delta_2 = \sum e_{x_i, y_j} & e_{x_i, y_j} \in \Delta_2 \end{cases} \quad (1)$$

If $\delta_1 > \delta_2$, edge set Δ_1 is called main order, and Δ_2 is called Against order, that is, Δ_2 is the back-tracking edge set.

Shots usually describe an event or part of the whole events and the development of event always have a time order, so the frame sets that describe the main contents of shots should also be an ordered set. The definition of back-tracking is used on the result of optimal matching to remove the back-tracking edge sets, and then get the timing constrained matching result. First, finding the main order of the key-frames set. Second, checking the back-tracking edges on the key-frames set, if exists, remove the back-tracking edges and find out the single point in X , then selecting point from Y with sampling intervals to match with the single points in X . Otherwise, the result of Kuhn-Munkres algorithm is the final optimal matching result. This idea is called amended optimal matching result, the detail are as the following.

Algorithm 1. Amended optimal matching algorithm

Step 1. Initialize Δ_1 and Δ_2 , if $x_i \leq y_j$, let $e_{x_i, y_j} \in \Delta_1$. Otherwise, $e_{x_i, y_j} \in \Delta_2$.

Step 2. If $\Delta_1 = \Phi$ or $\Delta_2 = \Phi$, then the algorithm is end. Otherwise, go to next.

Step 3. Using formula (1) to calculate δ_1 and δ_2 .

Step 4. If $\delta_1 \leq \delta_2$, find an edge e_{x_i, y_j} from Δ_1 . Suppose n_1 is the edge number of Δ_1 , divide the Y extent corresponding to x_i into $n_1 + 1$ part, take the first frame $y_{j_1}^{i_1}$ as the matching frame to x_i , $Y' = Y \cup y_{j_1}^{i_1}$, $n_1 = n_1 - 1$. If $n_1 = 0$, go to Step 6. Otherwise, go to Step 4.

Step 5. If $\delta_1 \leq \delta_2$, find an edge e_{x_i, y_j} from Δ_2 . Suppose n_2 is the edge number of Δ_2 , divide the Y extent corresponding to x_i into $n_2 + 1$ part, take the first frame $y_{j_2}^{i_1}$ as the matching frame to x_i , $Y' = Y \cup y_{j_2}^{i_1}$, $n_2 = n_2 - 1$. If $n_2 = 0$, go to Step 6. Otherwise, go to Step 5.

Step 6. Let $\Delta = \Delta_1 \cup \Delta_2$ and $M' = (X, \Delta, Y')$. Algorithm is end.

Using algorithm 1 to modify the Kuhn-Munkres optical matching and the matching result is shown in Fig.3.

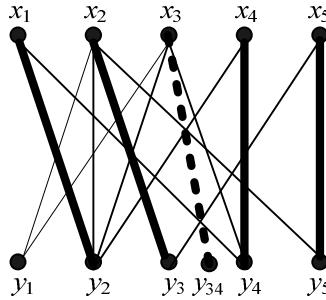


Fig. 3. Result of amended optical matching

In Fig.3, remove the optical matching algorithms produced the back-tracking edge e_{x_3, y_3} , the amended optical matching insert a new point y_{34} between y_3 and y_4 , let x_3 match with y_{34} . The amended algorithm eliminates back-tracking edges and the matching points corresponding by time sequence. Highlights the time series feature of video content and ensure the corresponding contents of video shot.

2.2 Similarity of Shots

According to the former analysis, calculates the similarity between shots Q and D with the following formula (2).

$$Shot_Sim(Q, D) = \frac{1}{n} \sum_{i=1}^n Sim(q_i, d_i) \tag{2}$$

n is couple number of similar key-frames and q_i, d_i is the couple key-frame of amended optical matching. Each similarity of key-frame in the whole shot is rather, so the mean value for the similarity is looked as the similarity of shot. If user can specify the importance of every couple key-frame for the shot before, and can set a weight for each couple of key-frame. At this time, the similarity of the whole shot can be calculated by the following formula.

$$Shot_Sim(Q, D) = \sum_{i=1}^n \bar{w}_i Sim(q_i, d_i) . \tag{3}$$

\bar{w}_i is the weight of similarity.

2.3 Experimental Results Analysis

The test video data, play during 90 minutes, including 237 shots, video frames number is 118500, is a daily operations training of an engineering team. The

configuration of computer for test the algorithm is CPU Pentium IV 2.4GHz and the size of memory 1 GB. Software platform is an algorithm verification system which is developed by Borland C++ Builder 6.0 and the database is oracle 8.0.5.

Video shot is described by key-frames which are got by the following ways.

- a) Sampling intervals (first frame, middle frame and the last frame of shot), method A.
- b) Clustering, method B.
- c) The article supposed analogy fuzzy C means clustering, method C.

Three kinds of shot are picked up from video database for testing.

Shot I, camera still, the commander assigned tasks only with the mouth and the whole scene is almost static.

Shot II, camera has three operating points, and the camera start from the first operation point and end to it. There has one sweep operation in the shot.

Shot III, camera still, equipment and operators enter the scene in order and then leave off the scene.

The three experiments using color and texture feature vectors of video key-frame image as retrieval features. Color characteristic of similarity is calculated with histogram intersection, the block policy introduce spatial distribution of key-frames and greatly reduce zero value appears when using original histogram. Then the two color features in the distance field and the similarity between them is proportional. Texture features are described with the contrast, energy and entropy of the GLCM, and using the mean and standard deviation, μ_{CON} , μ_{ASM} , μ_{ENT} , μ_{COR} , σ_{CON} , σ_{ASM} , σ_{ENT} , σ_{COR} , as components of those texture features. The final similarity can better reflect the two frames in color and texture and the experimental results proved the advantages of the proposed algorithm from the similarity measure of shots. A part of experimental results is showed in Table 1.

Table 1. Retrieval results of the three methods

Shot	Method A			Method B			Method C		
	R(%)	P(%)	S(s)	R(%)	P(%)	S(s)	R(%)	P(%)	S(s)
Shot I	100	93	9.000	100	98	3.000	100	100	3.000
Shot II	58	87	9.000	95	85	8.974	100	97	11.072
Shot III	73	55	9.000	98	73	11.857	100	100	15.483
Average	77	78.3	9.000	97.7	85.3	7.944	100	99	9.852

R is Recall, P is Precision and S is retrieval response time.

The table 1 shows that the three methods for the type of shot I have high recall and precision. The retrieval response time of method A is significantly less than method B and C for the number of key-frames. Retrieval result of method A is basically no contact with the content of shot, no matter what changes on the main contents of shot, the algorithm is basically the same response time of retrieval. The core of method b

and *c* are based on clustering to extract key-frame and the result sets of key-frames may be the same when the content of shot is none self-similar. Method C will extract greater than method B in key-frames when the content of shot is self-similar, because method C considers the video sequence in clustering and method B put the self-similar content to the same cluster and the final number of key-frames must be less than method C. Therefore, the latter two methods in the recall rate are roughly equal. Method C describe the main content of shot in detail, for it takes time sequence of video into account when extracting the key-frame, so the precision obtained relatively higher than the method B. Method *c* extracts the number of key-frames to be more than method B. However, in the maximum matching algorithm that improved, which requires the matching of key frames can not exist back-tracking and reduces the search time of the key-frame pairing. So the method C has relatively longer response time.

The statistical results of table 1 show, in terms of retrieval performance for the algorithm, the recall of C method was 100% and higher than 97.7% of method B. In accuracy rate, the recall of method C was 99% and exceeds nearly 14 percentage points than method B. Therefore, the performance of the proposed method is better than clustering.

3 Conclusion

The advantage of this algorithm is the in introduction of the video sequence characteristics during the optimal matching. More accurately capture the details of the video content changes, so as to achieve effectively improve the recall and precision of shot retrieval results. In addition, during the maximum matching, there were no deliberate required two shots to be matched with the same number of key-frames, but to be retrieved base on the query shot. The similarity is the maximum matching of shots when all key-frames in query shot are matched successfully. On the other hand, insert key-frames with sampling intervals in the corresponding positions of matching shots to match with the query shot.

References

1. Yuxin, P., Chong-Wah, N., Jianguo, X.: An Approach for Shot Retrieval by Optimal Matching in the Bipartite Graph. *Acta Electronica Sinica* 32(7), 1135–1139 (2004)
2. Yuxin, P., Chong-Wah, N., Qingjie, D., et al.: An Approach for Video Retrieval by Video Clip. *Journal of Software* 14(8), 1409–1417 (2003)
3. Huiquan, S.: *Graphic theory and application*. Sciences Press, Beijing (2004)
4. Yujin, Z.: *Content based visual information retrieval*. Sciences Press, Beijing (2003)
5. Dongru, Z.: *Video Introduction to Database Management System*. Sciences Press, Beijing (2002)
6. Li, D., Lizuo, J., Shumin, F.: Ensemble Similarity-Blased Video Retrieval. *Journal of Electronics & Information Technology* 29(5), 1023–1026 (2007)

Improved Cue Fusion for Object Tracking Algorithm Based on Particle Filter

Hui Li and Li Zhang

School of Electrical Engineering and Automation, Henan Polytechnic University,
Henan 454010, China
li20042007@163.com,
zlfenger@163.com

Abstract. The traditional object tracking with cue fusion is inaccurate under complex background. Especially when some blocks exist, the targets may be lost. To solve this problem, improved cue fusion for object tracking algorithm based on particle filter is proposed. It uses color and motion as the observation information source. Color is the main observation information and motion is the auxiliary information. It weights particles followed by the order of information. Block detection, particle filter and mean-shift are used together to track the interest targets. The experimental results show that in complex scene, when the number of particles of the proposed method is half of the traditional cue fusion, the proposed method can improve effectively the accuracy of target tracking, and track object stably when the shape is changing. So the proposed method is more robust and real-time.

Keywords: Object Tracking, Cue Fusion, Particle Filter, Mean-shift.

1 Introduction

In recent years, the application and development of video tracking is rapid. Particle filter is a common method for target tracking, because of its random walk and capability of nonlinear [1-4]. Mean-shift algorithm has many applications in real-time tracking for its iteration with no parameters [5-7]. We can combine particle filter and mean-shift to realize real-time tracking of anti-blocking by using their advantages, while the degradation of particle filter may lead iteration to local extremum, and lose the target [8]. Literature [9] proposed that the moving target is blocked to deal with occlusion, but the method is not enough stability in a complex environment, and every block is not given effective weight.

Cue fusion is necessary in target tracking. It can get rid of the disadvantage of single source, bring better stability for tracking performance. In this paper, we combine the characteristics and performance of these algorithms, using stratified sampling strategy of cue fusion [10], detecting sub-block, to tracking targets of interest. The result show that the method can track the target accurately and reliably, and handle the situation of target blocked.

2 Particle Filter and Mean-Shift

2.1 Mean-Shift

The main ideal of mean-shift is that we find of the maximum of probability density in the neighborhood by iteration. It is a high efficiency, non-parameter algorithm. The iterative equation from the current location y_0 to the next frame y is:

$$y = \frac{\sum_{k=1}^N x_k s_k f\left(\left\|\frac{x_k - y_0}{h}\right\|^2\right)}{\sum_{k=1}^N s_k f\left(\left\|\frac{x_k - y_0}{h}\right\|^2\right)} \quad (1)$$

Where x_k is the coordinates of k th pixel in the target area, s_k is corresponding weights, N is the total number of pixels, $f(x)$ is the kernel function, h is the bandwidth of the window and is half of width of window [11]. The result of Iteration is that $\|y - y_0\|$ is less than the threshold.

2.2 Particle Filter

Particle filter has good performance in non-Gaussian and nonlinear environment. Its main ideal is that according to sampling enough samples, it approximates the posterior probability distribution, and then estimates the true value of the system. Samples can be simulated by a set of particles and its weight $(x_t^{(i)}, \omega_t^{(i)})$. The basic particle filter is:

- (1) Suppose the particle of time $t-1$ is $(x_{t-1}^{(i)}, \omega_{t-1}^{(i)})$, $i = 1, 2, 3 \dots N$.
- (2) Prediction by one step: $X_t = X_{t-1} + v_t$ Where v_t is additive white Gaussian noise with zero;
- (3) Weight calculation:

$$\omega_t^{(i)} \propto \omega_t^{(i-1)} \times \frac{p(y_t | x_t^{(i)}) p(x_t^{(i)} | x_{t-1}^{(i)})}{q(x_t^{(i)} | x_{t-1}^{(i)}, y_t)}, \quad \omega_t^{(i)} = \frac{\omega_t^{(i)}}{\sum_{i=1}^N \omega_t^{(i)}}; \quad (2)$$

For calculating easily, importance function is:

$$q(x_t^{(i)} | y_{t-1}^{(i)}, y_t) = p(x_t^{(i)} | x_{t-1}^{(i)}) \quad (3)$$

- (4) Resampling;
- (5) Calculating the optimal value of the state \hat{x}_t

$$\hat{x}_t = \sum_{i=1}^N x_t^{(i)} \times \omega_t^{(i)} \quad (4)$$

In order to avoid degradation of particle, resampling is used to slow down trend. While the iteration of mean-shift may produce "optimized" and make the particle in maxima of source of interference, when occlusion or other disturbance exists. In this paper, we solve the problem of "optimized" and degradation of particle by using detecting sub-block and the idea of [12].

3 Object Tracking with Cue Fusion Based on Particle Filter

3.1 Color Model

In order to improve the reliability of the histogram, we use the weighted color histogram to create a color model. The color space is divided into m sub-interval, the total number of pixels in region is M , and every pixel is x_k . The weighted color histogram is $q = \{q_u\}_{u=1,2,\dots,m}$.

$$q_u = C \sum_{k=1}^m f \left(\left\| \frac{x_k - y}{h} \right\|^2 \right) \delta[b(x_k) - u] \tag{5}$$

Where C is Normalized constant, makes sure $\sum_{u=1}^m q_u = 1$. $\delta(x)$ is Kronecker function, $b(x_k)$ is corresponding characteristics of the current pixel, y is the center coordinates of current region.

The degree of similarity candidate region and target template can be measured by Bhattacharyya similarity coefficient, when coefficient is bigger, they are more similar each other.

$$\rho(p, q) = \sum_{u=1}^m \sqrt{p_u(y)q_u(y_0)} \tag{6}$$

Finally, the particle is weighted by color likelihood function. The function is defined as follows:

$$P(p, q) \propto \exp(-D^2(p, q) / 2\sigma_c^2) \tag{7}$$

$$D^2(p, q) = \sqrt{1 - \sum_{u=1}^m \sqrt{p_u q_u}} \tag{8}$$

Where is the standard deviation of the color likelihood function.

3.2 Motion Model

Motion model is created by motion likelihood function according to literature [10]. The motion information is robust, when light is changing. In order to calculate simply, difference image can be extracted through two frame difference. Changes of

movement can cause spread of gray values, so gray histogram with the uniform distribution is used.

$$h_{u,q} = \frac{1}{m} \quad u = 1, 2, \dots, m \quad (9)$$

The particle is weighted preliminarily by motion likelihood function. The function is defined as follows:

$$P(y) \propto \exp(-D^2(h_p, h_q) / 2\sigma_m^2) \quad (10)$$

Where σ_m is the standard deviation of the motion likelihood function.

3.3 The Realization of Fusion Algorithm

The proposed algorithm uses the stratified sampling strategy [10]. According to the information observed in the proportion of state estimation, the order is determined to make algorithm more accurate. Because of its characteristics of permanent and strong anti-blocking, the color information is main source of information, while movement information is preliminary judge information. Algorithm procedure is as follows:

- (1) Tracking target is divided into two sub-blocks 1,2 and then sampling by equation (3) to obtain the initial particle $x_{1,t}^k, x_{2,t}^k$;
- (2) Weighting the initial particle for two sub-blocks;
- (3) Resampling;
- (4) Reading into the target frame image, establishing the weighted color histogram by equation (5), and using mean-shift to calculate the location of the current frame for the two sub-blocks;
- (5) Calculating the similarity of target area and the template for two sub-blocks;
 - a: If the similarity of two sub-blocks are less than preset threshold go to 6).
 - b: If one is less than the threshold, there is some shelter, and the current sub-block is the main object. The other is dropped, go to 6).
 - c: If two sub-blocks is both greater than the threshold, it is completely blocked. Iterating the current coordinates, go to 7);

4 Simulation Results and Analysis

In this paper, two video in two different environments is simulated to validate the performance of the algorithm. Color pace is 10×10×10 RGB, image resolution is 320×240.

Figure 1 shows the results about algorithm of this paper and literature [10]. In comparison, the number of particles is 50 in this algorithm, while the number of particles is 100 in literature [10]. It can be seen from the figure, when there is completely blocked in the vehicle or part of shelter, the literature [10] algorithm can only track the target, but not accurate, the improved algorithm can accurately track the vehicle by using half particles, because the good performance of mean-shift when

detecting sub-block make algorithm more precise. Particularly, because the car suddenly accelerated in 53 frame, literature [10] algorithm can only track the target, due to the good characteristics of mean-shift algorithm, it makes the proposed algorithm are more accurate on tracking target.

Figure 2 shows that in complex traffic environment, the cue fusion and good drift performance of mean-shift and can still track the target accurately when the target of interest is very small. The 58 frame shows that the proposed algorithm can judge the location of target. The 112 shows the accuracy of algorithm when the target is very small.

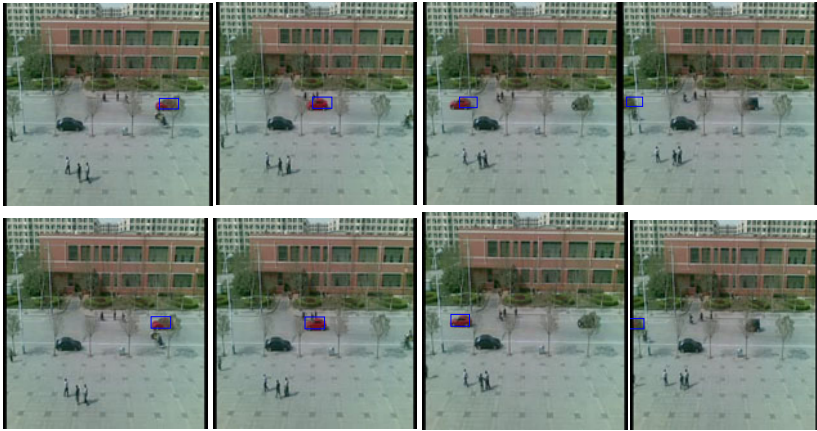


Fig. 1. The Tacking Results in Trunk Road of a School (the line of 1: the algorithm of literature [10]; the line of 2: the algorithm of this paper. Frames are 11, 33, 53, and 64)



Fig. 2. The Tacking Results in High-way with Intensive Vehicle (Frames are 20, 58, 85, and 112)

5 Conclusion

An improved object tracking with cue fusion based on particle filter is proposed in this paper. On the one hand, the combination of detection in sub-block and mean-shift can effectively improve the problem of accuracy under blocked; on the other hand, it can track the target interested accurately in complex traffic environment, because of

its permanent characteristics of color and features of mean-shift algorithm. The next step is to study multi-target tracking in the complex scenario environment, and the stability of multi-target tracking.

Acknowledgement. This work is supported by the Natural Science Foundation of China (Grant No 41074090), Henan Province Open Laboratory for Control Engineering Key Disciplines (Grant No KG2009-18) and Doctorate Program of Henan Polytechnic University (Grant No B2009-27).

References

1. Ying, Z.G., Pietikainen, M., Koller, D.: Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE. T. Pattern. Anal.* 29, 915–928 (2007)
2. Pan, P., Schonfeld, D.: Visual tracking using high-order particle filtering. *IEEE. Signal. Proc. Let.* 18, 51–54 (2011)
3. Bouaynaya, N., Schonfeld, D.: On the Optimality of Motion-Based Particle Filtering. *IEEE. T. Circ. Syst. Vid.* 19, 1068–1072 (2009)
4. Bolic, M., Djuric, M.P., Hong, S.: Resampling algorithms and architectures for distributed particle filters. *IEEE. T. Signal. Proces.* 53, 2442–2450 (2005)
5. Ilonen, J., Kamarainen, K.J., Paalanen, P., Hamouz, M., Kittler, J., Kalviainen, H.: Image feature localization by multiple hypothesis testing of Gabor features. *IEEE. T. Image. Process.* 17, 311–325 (2008)
6. Huang, K., Aviente, S.: Wavelet feature selection for image classification. *IEEE. T. Image. Process.* 17, 1709–1719 (2008)
7. Han, R., Jing, Z., Li, Y.: Kernel based visual tracking with variant spatial resolution model. *Electronics. Lett.* 44, 517–518 (2008)
8. Jia, Y., Yuan, W.M., Zhen, C.S., Lin, Z.Q.: Anti-occlusion tracking algorithm based on Mean Shift and fragment. *J. Opt. Precision. Eng.* 18, 1413–1419 (2010)
9. Nejhum, M.S., Ho, J., Yang, M.: Visual tracking with histograms and articulating blocks. In: 26th IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1–8. IEEE Computer Society, Alaska (2008)
10. Hua, H.Z., Liang, S.Y., Qaun, L.D., Xin, F.: A particle filter based tracking algorithm with cue fusion under complex background. *Optoelectronics-Laser* 19, 678–680 (2008)
11. Comaniciu, D., Ramesh, V., Meet, P.: Real-time tracking of non-rigid objects using mean shift. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 142–149. IEEE Computer Society, Hilton Head (2000)
12. Liang, C.F., Li, M., Xiao, L.Z., Zheng, Q.Y.: Target Tracking Based on Adaptive Particle Filter Under Complex Background. *Acta. Electr.* 34, 2150–2153 (2006)

A Study on Bus Routing Problem: An Ant Colony Optimization Algorithm Approach*

Min Huang

Department of Electronic and Information Engineering,
Qiongzhou University, Sanya 572200, China
Huangmin198100@126.com

Abstract. Reducing the transportation time on bus routes can increase the proportion of regular bus passengers, thus easing the traffic congestion. This paper presents a new model of bus routing based on an ant colony optimization (generally abbreviated as ACO) algorithm. The model, comprehensively taking traffic variables into account, figures out the quickest and optimal route, and is proved to be feasible by conducting researches and simulation experiments.

Keywords: ACO algorithm, Traffic congestion, Density of Traffic stream, Optimal route.

1 Introduction

Traffic congestion has been one of the biggest problems in urban transportation. To relieve the traffic congestion and pressure, metropolitan governments advocate the use of public transport [1]. Since passengers incline to choose the quickest route when setting out, reduction of bus trip time will increase the proportion of regular bus passengers. Therefore, a study of bus routing problem is of great significance and practical value.

The ACO algorithm is a bionic algorithm mimicking the collective exploitation of food sources among animal societies. In the practice of best-route searching, the ACO algorithm can give dynamic feedbacks to outer influence, and is highly feasible and adaptable to bus route optimization. This paper presents an approach for bus routing based on an ACO algorithm .

2 The Principle of ACO Algorithm

The Ant System was first proposed by Italian researchers M. Dorigo and V. Maniezzo in 1991 [2]. After years of study biologists have observed that ants, though without

* This work was supported by Scientific Research Foundation of Qiongzhou University (No. QYQN201138) and Research project of Institution of Higher education of Hainan Education Department of China (No. Hjkj2010-40) and Project of technological collaboration of the college and local government of Sanya city of China (No. 2010YD28).

sense of sight, laid down a special substance---pheromone trails on the way they traversed. Ants, when facing an intersection, wander randomly down one of the paths, depositing a trail of pheromone detailing the path's length. The longer the route is, the more pheromones will evaporate [3]. When other ants come to the same intersection, they will be inclined to follow the track with thicker pheromones, hence creating a positive feedback mechanism. The optimal path will be increasingly enhanced whereas other routes will eventually disappear as pheromones evaporate with time. Therefore, the best route will eventually be figured out by the colony of ants. It is thus clear that an individual ant's cognitive ability is limited, but by depositing pheromone trails to exchange information, the ant colony can collectively work out an optimal solution [4].

3 Bus Route Optimization Based on ACO Algorithm

3.1 Definition of Optimal Path

Given a list of certain bus stops, there are several different possible routes stringing them up. Traffic condition, density and flow vary at different moments on different roadway segments, resulting in varied travel time of each route. On routes of roughly equivalent length, a bus takes more time on a congested route with higher traffic density than on others. Therefore, with a given amount of bus stops, a route with better traffic condition, lower density and smoother flow as well as shorter length is ideal for passengers, and is the optimal path [5].

3.2 Description of the Routing Model Based on ACO Algorithm

Bus routing problem is similar to the behavior of ants searching for food. In a given urban transportation network with n points, the starting point A is the ant nest, and the terminal point B is the food source. An individual sets out from point A , visits certain roadway segments and points, and arrive at point B in the end. In the phase of routing, the individual chooses a suitable route based on a heuristic combining the dynamically changing conditions of both inner and outside world. In this way, the bus routing problem can be modeled as an ant colony optimization problem.

Suppose that the population of ants (size of ant colony) is m , N is the number of points, T represents the times of cycles, d_{ij} ($i, j = 1, 2, \dots, n$) is the distance between point i to point j . η_{ij} is computed by some heuristic indicating the a priori desirability of point transition i,j . And $\eta_{ij}(0) = e$, where e is a constant. $\tau_{ij}(t)$ represents the amount of pheromone deposited on edge (i,j) at moment t ($0 < t < T$). A and B respectively represent the starting point and the destination. At the beginning, all ants assemble at point A , and the amount of pheromone deposited on every edge is equivalent, namely $\tau_{ij}(0) = C$ (C is a constant). Each ant chooses the next point to visit based on the amount of pheromone deposited on the edge. The probability $p_{ij}^k(t)$ indicates the probability of ant k moving from point i to point j at moment t . In general, ant k moves from current point to the next one with probability

$$P_{ij}^k(t) = \begin{cases} \frac{\tau_{ij}^\alpha(t)\eta_{ij}^\beta(t)}{\sum_{\alpha \in allowed_k} \tau_{ia}^\alpha(t)\eta_{ij}^\beta(t)} & j \in allowed_k \\ 0 & otherwise \end{cases} \tag{1}$$

where point j is the next targeted point that ant k is heading for, $tabu_k (k = 1, 2, \dots, m)$ records points having been traversed by ant K , $allowed_k$ is the next point allowed to be chosen. α is the heuristic factor of pheromone, β is the expected heuristic factor [6].

$\eta_{ij}(t)$ is a heuristic function. Typically, $\eta_{ij}(t) = \frac{1}{d_{ij}}$, indicating the desirability of point transition from i to j . But it only details the distance between the current point and the next one. Thus, a new parameter D_{ij} is added. D_{ij} indicates the traffic condition, namely density of the traffic stream, at roadway segment i to j at different moments. Then, the typical function is updated to be $\eta_{ij} = \frac{1}{\alpha_1 d_{ij} + \alpha_2 D_{ij}}$, where

$\alpha_1 + \alpha_2 = 1$. α_1 and α_2 are both parameters indicating how much the edge length and travel time are weighted for the solution. If $\alpha_1 > \alpha_2$, the edge length is weighted. If $\alpha_1 < \alpha_2$, the travel time is highlighted. If $\alpha_2 = 0$,

$\eta_{ij} = \frac{1}{\alpha_1 d_{ij} + \alpha_2 D_{ij}}$ then equals to $\eta_{ij}(t) = \frac{1}{d_{ij}}$. And $D_{ij} = \delta(t)$, $\delta(t)$ is the density of traffic stream, an traffic condition indicator, on roadway segment i to j at a certain moment. D_{ij} , α_1 and α_2 enable buses to better choose the optimal route at different time segments.

When ant k has completed a solution, the pheromone information deposited on the edge it crossed will be updated according to the local trail updating rule. The local pheromone will be:

$$\tau(i, j) \leftarrow \rho\tau(i, j) + \Delta\tau^k(i, j) \tag{2}$$

$$\Delta\tau^k(i, j) = \begin{cases} \frac{Q}{L_k} & \text{If ant } K \text{ uses curve(} i, j \text{) in its tour,} \\ 0 & otherwise \end{cases} \tag{3}$$

where Q is a constant, ρ is the pheromone persistence coefficient, L_k is the total length of ant K 's tour. When ants have all completed a tour, the global optimal edge will emerge according to the global trail updating rule. The global pheromone will be updated to:

$$\tau(i, j) \leftarrow \tau(i, j) + \omega * \Delta\tau(i, j) \tag{4}$$

$$\Delta\tau(i, j) = \begin{cases} \frac{1}{d_g} & \text{If edge (i, j) is the global optimal solution} \\ 0 & \text{If edge (i, j) is not the global optimal solution} \end{cases} \tag{5}$$

where d_g is the global optimal route, ω is the pheromone evaporation coefficient affecting the convergence speed.

4 Algorithm Model Simulation

4.1 Test

In order to prove the feasible of the algorithm, the paper carry through a simulation test. In transportation, the relationship between the vehicle speed V and the traffic density K is inverse proportion, whose function relates to the practical transportation. Pipes-Mun jal provided a group of series of curves to describe the relationship between them:

$$V = v_f (1 - K / K_j)^n \tag{6}$$

where v_f refers to drag-free speed; K_j refers to jam density; n refers to parameter, whose value is confirmed by the section transportation data in investment. Commonly $n=1$,but when the traffic density is heavier, $n<1$; when it becomes lighter, $n>1$; when $1 - K / K_j < 0$, $V=0$ [7].As to a section, its traffic speed in different sections can be figured out through (6),and thus we can get the average speed and the time of using in this section. In the algorithm, we choose the appropriate parameters to carry out a great deal of test, the parameters are as follow: $\alpha = 1.0, \beta = 1.5, \rho = 0.7, Q = 10$, since the test focuses on the problem of saving time or not, so $\alpha_1 = 0.4, \alpha_2 = 0.6$. We choose a time section in city A, and time 1, we make an investment for the route of section A in transportation, and make use of the traffic density ect. To do the simulation test.

4.2 The Comparison the Test Result

By doing the simulation test between the traditional ant colony algorithm and the improved ant colony algorithm that proposed by the paper, we get the results as shown in Table 1, Fig 1 and Fig2.

Table 1. The Experimental Result

time	Tradition al ACO algorithm	Improved ACO algorithm		Traditional ACO algorithm	Improved ACO algorithm	
	Average speed	Average speed	D- value of them	Time of using	Time of using	D- value of them
1	171.25	246.25	75	15.3	11.7	3.6
2	297.5	297.5	0	8.8	8.8	0
3	175.01	332.52	157.51	15	8	7
4	187.4	273.76	86.36	14	10	4
5	271.26	293.77	22.51	10	8.9	1.1

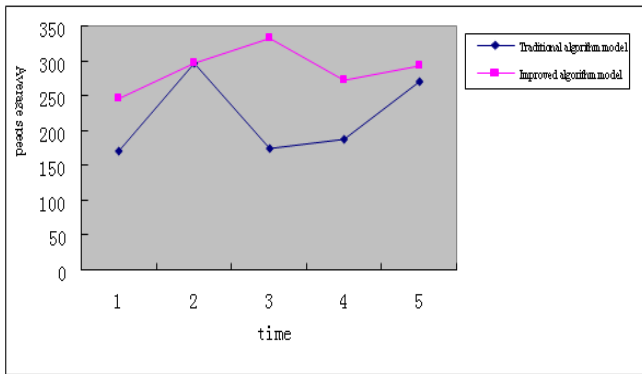


Fig. 1. Average Speed Compare Among Traditional Algorithm Model and Improved Algorithm Model

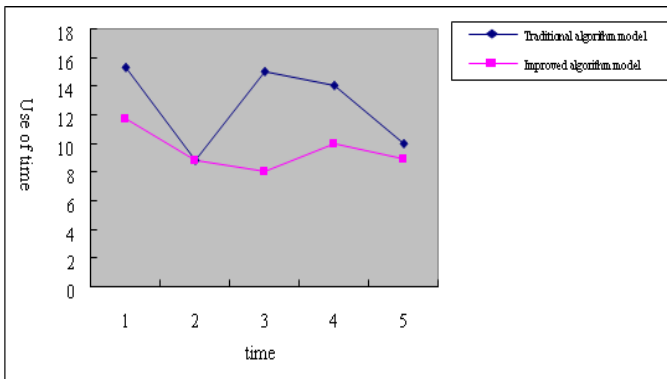


Fig. 2. Use of Time Compare Among Traditional Algorithm Model and Improved Algorithm Model

5 Conclusion

The paper tries to add the transportation factor: traffic density to ant colony algorithm in order to solve the problem of selecting the best routes in public transportation. Numerical simulation test results show that the algorithm is feasible. The problem of transportation is a quite complex dynamic problem, thus the next step of the research should be get further research on the algorithm.

References

1. Xu, X.: On the priority to public traffic in city traffic. *Technology and Economy in Areas of Communications* 2, 67–68 (2005)
2. Ketcha, V., Ann, N.: Colonies of learning automata. *IEEE Transactions on Systems, Man, and Cybernetics Part B* 32(6), 772–780 (2002)
3. Haibin, D.: *The Theory of Ant Colony Algorithm and Its Application*, pp. 8–20. Science press, Beijing (2005)
4. Yi, Z., Liang, Y.: The Improve Ant Colony Algorithm Based on the optimization of selecting routes. *Computer Engineering and Application* 43(13), 22 (2007)
5. Huang, G., Gao, X., Tan, F., et al.: On the Selection of Optimal Path and its Application Based on Ant Colony Algorithm. *Computer Engineering and Application* 43(13), 233–235 (2007)
6. Dotigo, M., Maniezzo, V., Colomi, A.: Ant system: optimization by a colony of cooperating agents. *IEEE Transaction on Systems, Man, and Cybernetics-Part B* 26(1), 29–41 (1996)
7. Yang, Y., Liu, X., Yu, Q., Chu, S.: Research on Three Traffic Flow Parameters. *Journal of Beijing University of Technology* 32(1), 43–45 (2006)

The Research of ANN Forecasting Mode Based on Cloud Platforms*

Jiang Xuesong**, Wei Xiumei, Geng Yushui, and Wang Xingang

Shandong Polytechnic University, School of information,
250353, ShanDong Jinan, China
jxs@spu.edu.cn

Abstract. With the rapid development of Cloud computing, data mining and ANN technology, we try to combined cloud data mining and ANN technologies, Using the Capacity of these technologies such as data acquisition, data storage and analyzing large amounts of complicated non-linear data, To construct a forecasting model for massive data real-time accurately which is difficult to complete on the traditional mode. This new model utilizes an enterprise sales data to test. The result proves that the model is feasible and effective. According to this model's prediction result, we can provide scientific and real-time decision support for the enterprise management.

Keywords: cloud computing, Ann, RBF, forecasting, data mining.

1 Introduction

Along with the development of computer, internet, communication technologies, and the automation of operational processes in all, almost all business generates tens of hundreds of GB or even TB of historical date[1]. How to forecast based on the massive data? The traditional forecast systems can not meet the new requirements any more, which is mainly manifested in the following points: data forecasting involved the dealing with massive data, so the traditional forecast systems can not satisfied the requirements of Operating efficiency, Computing performance, accuracy and storage space; the data have much incompleteness, noise and inconsistency which leads forecasting to chaos[2]. In this case, we should use data mining and cloud platform to store and compute the massive data. Data mining is the process of computers extracting implicit unknown information or mode with potential applications [5].

Cloud computing is the development of distributed computing, Parallel Computing and grid Computing [3]. It is a resource pool with a large number of Virtual Resources such as Development Platform, hardware and I/O services. Depending on the load, these virtual resources can be dynamically reconfigured, so that it can improve resource utilization. Cloud storage is achieved by using distributed storage to ensure

* Fund project: Development of information industry of ShanDong province special fund project (2008R00041).

** Corresponding author.

high reliability, availability and economy[4]. Therefore cloud computing system can satisfy the requirement of massive data storage, operation and computing more cheaply. This system can provide data forecasting based on massive data services for users.

Now popular forecasting methods are as following: Statistical methods based on probabilistic model, Decision tree based on machine learning and ANN and so on.

ANN methods are robust, Self-organizational, adaptive, parallel processing. ANN can analyse a large number of complex data, and exact extremely complex modes. So ANN methods are very suitable to resolve the data forecasting.

2 Construction of ANN Forecasting Model

ANN forecasting model based on cloud platform is shown as figure 1:

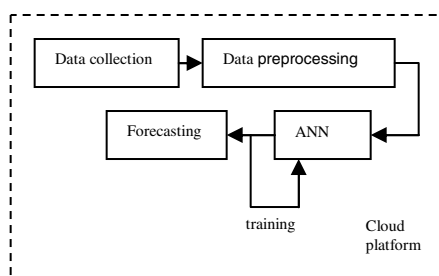


Fig. 1. ANN forecasting model based on cloud platform

Forecasting model based on RBF neural networks and BP neural networks are the most commonly used methods. Both RBF and BP neural networks are mature technology. So they are used at data forecasting successfully. Both RBF and BP neural networks are feed forward artificial neural network. There are some similarities and differences between them.[6][7].

3 The Result of Experiment Verify the Forecasting Effect

This paper take the various manufacturing enterprise product sales forecast as an example to verify the effect of using BP and RBF for data forecasting.

Sales forecast is one of the most important factors in the enterprise marketing management. And it also is the key link of enterprise supply chain.[1]

Based on the analysis and research affecting the market supply and demand of the many factors and the present and past sales data, sales forecast uses scientific method to estimate and speculate the future market products supply and demand development trend. According to the sales forecast results, the enterprise may set up a rational purchasing plan, production plan, stock plan and marketing plan.

3.1 Evaluation Standard of Data Forecasting

To compare the BP model and RBF model predictive power, we adopted mean absolute error percentage MAPE to evaluate accuracy of prediction.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - f_i}{y_i} \right| \times 100\% \quad (1)$$

In the expressions (1), y_i is the actual output of the i node, f_i is the prediction of the i node.[2]

3.2 Data Collection

Collecting the data from web, we need to use web data mining technology. Web data mining uses data mining technology to find and retrieve information from web documents and services. Because there are lots of information available on Internet, we need distributed operation in clouds platform.

Considering the various factors affecting the product sales, we choose the sum of investments in fixed assets used, producer price index, output of major industrial products, purchasing price indices of raw material, fuel and power, the related industrial index, gross domestic product, price of production sales, quality of products, type of production etc. 11 main influencing factors as the input of the model.

So the number of input cells is 11. The output of the model is the forecasted sales after neural network model calculation. The number of output cells is 1.

The sample data of this model is from the National Bureau of Statistics of China website. And the products data is from a glass fiber manufacturing enterprise's production and sales statistics in 2006, 2007, and 2008. Data of 2006 and 2007 is the training data. After training the algorithm and adjusting parameters, we determine the final neural network model. Data of 2008 is the test data to provide the forecasting effect of the ANN model.

3.3 Data Pretreatment

Because we collect the mass of data, and some of them are useless or damaged information, we must preprocess the data before analysis. How to preprocessing the data?

We should classify, screen, and preserve them in a standard format. So that it is easy for us to input analyze module. After preprocessing the data, we can get a set input cells (x_1, x_2, \dots, x_n) to describe the users' behavior characteristics.

Because collecting data is not in the same order of magnitude, they must be mapped to [0,1] interval. After pretreatment the data are normalized to neural network's input.

Because the data for pretreatment of is numerous, and normalized calculation is complex, we should use clouds platform to calculate to improve computational speed and efficiency. [6][7]

Take X_{ij} for the training data set, which contains a large amount of data element changes, the goal is to make the change between elements less than 1. First calculate the average of all elements X :

$$p_0 = \frac{1}{m * n} \sum_{i=1}^m \sum_{j=1}^n x_{ij} \quad (2)$$

In the expressions (2), $i=1,2, \dots, m$; $j=1,2, \dots, n$.

And then calculate the standard deviation of all the elements of X:

$$p_1 = \sigma(x_{ij}) \quad (3)$$

In the expressions (3), σ --Standard deviation.

So the normalized X is:

$$x_{ij} = \frac{x_{ij} - p_0}{p_1} \quad (4)$$

The normalized data as the neural network model of the input data, the data of 2006, 2007 and 2008 are stored in three tables.

Table 1. Normalized data

No	x1	x2	x3	...	x11	Actual sales
1	0.16221	0.60714	0.4902	...	0.54098	0.15678
2	0.04097	0.57143	0.43137	...	0.38883	0.52347
3	0	0.5	0.2549	...	0.31325	0.23654
4	0.055184	0.35714	0.13725	...	0.06135	0.36983
...

The data in Table 1 is the result of normalization, so all data are in [0, 1] interval.

3.4 The Experimental Results of BP Network and RBF Network

In this paper we use matlab6.5 to train the BP network and RBF network. First, we train the BP network. After repeated testing, we select the number of hidden layer neurons is 25. BP network for each input and output sample value, each power affects the output of the network. During the training, each power needs to be adjusted through the back error propagation, thus learning is slowly.

In Table 1 take x1, x2, ... , x11 as input data of the BP network, to train the network, and then get the output data, ie the forecast to the sales. Compare the predicted values to actual sales in Table 1, adjust the parameters of the network repeatedly, then we will get the final model. The BP network took 7.2650 seconds, 120 times training to the goal error of 0.01, and then the training stops.[7]

But when we use BP network, during the training, not converge condition often occurs.

The test sample set data of 2008, input to the trained BP network model to predict the effect of the model. After the BP network computing, we get the output data. Compare the predicted data to the actual sales data in 2008; we can see the final prediction results of the BP network.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{f}_i}{y_i} \times 100\% \right| = 2.22\% . \tag{5}$$

After training the RBF neural network model and adjusting the parameters, we get the ultimate model of RBF neural network. The RBF network took 0.254 seconds to the goal, and stop the training. not converge condition did not occur.

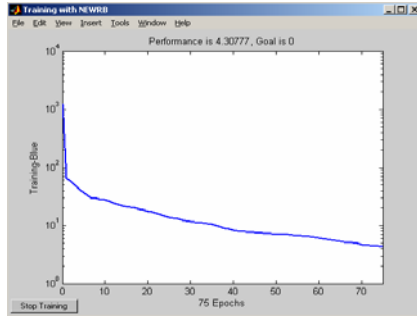


Fig. 2. The training of RBF network

From fig 2, we can see that the convergence speed is relatively fast.

Input the test sample set data of 2008 to the trained RBF network model to predict the effect of the model. After the RBF network computing, get the output data. Comparing the predicted data to the actual sales data in 2008, we can see the final prediction results of the RBF network.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{f}_i}{y_i} \times 100\% \right| = 0.82\% . \tag{6}$$

4 Conclusion

Through the experiment, it can be proved that RBF neural network has better forecasting effect than BP neural network. But the results of RBF are affected by various parameters, so the prediction accuracy will reduce with the passage of time. So the RBF network does not use on long-term forecast. We can predict after period of time, and then put the new data as the training data on network model to training, in order to achieve higher precision of prediction.

References

1. Bo, C., Zhu, O.: Prediction of winter wheat evapotranspiration based on BP neural networks. Transactions of the CSAE 26(4), 81–86 (2010)
2. Chang, F., Dean, J., Chema Wat, S.: BigTable: A distributed storage system for structured data. ACM Transactions on Computer System 26(2), 1–26 (2008)

3. Cai, X.-w., Duan, M.-b.: Research about Cloud Computing and Using at Data Mining. *Computer Knowledge and Technology* 6(22), 6272–6274 (2010)
4. Gui, B.-x., He, J.: Distributed Data Mining Approach with High Performance Cloud. *Computer Engineering* 36(5), 76–78 (2010)
5. Li, L.: Research of sale forecast based on data mining. Example of Application 8, 74–76 (2009)
6. Wei, W.: *Neural Networks compute*, pp. 41–46. Higher Education Press
7. Yan, P.-f., Zhang, C.-s.: *Artificial neural network and Simulated evolutionary computation*, pp. 46–48. Tsinghua University Press

Trend Analysis of 2011 China's Car Sales Volume and Carbon Emissions Based on Rough Set Theory

Li Yuansheng¹, Yang Yang^{2,*}, Fu Yanxiao¹, and Xu Xiangyang¹

¹ School of Transportation Science and Engineering, Beihang University, 37 Xue Yuan Road, Haidian District, Beijing 100191, China

² School of Transportation Science and Engineering, Beihang University, 37 Xue Yuan Road, Haidian District, Beijing 100191, China, +86 15801493886

lys@237auto.com, yy3572.beijing@yahoo.com.cn,
fyxbj2008@yahoo.com.cn, xxy@buaa.edu.cn

Abstract. With the fast growth of Chinese economy and the society's progress, the environment issues in China begin to attract people's attention more and more. After United Nations Climate Change Conference in Copenhagen, Chinese begin to concern about the "carbon emissions" and the low carbon lifestyle. According to the expert's forecast, China's automobile industry's carbon emissions possibly occupy about 5% of total emissions. And the automobile industry of our country is developing quickly, its carbon emissions rate in total emissions will be more and more. In this paper, it's demonstrated that Rough Set Theory can find some potential knowledge by data analysis. Based on Rough Set Theory and the demand of prediction of car sales volume, a procedure of car sales forecast is proposed. Then automobile carbon emissions can be forecasted with obtained rules set and calculation of carbon emissions. The prediction is expected to have some effect on the establishment of Chinese twelfth five-year plan, especially the policy about the carbon emissions.

Keywords: carbon emissions, low carbon, Rough Set Theory, car sales volume, forecast.

1 Introduction

After United Nations Climate Change Conference in Copenhagen, carbon emissions which is a product of fossil fuel (such as coal, oil and natural gas) consumption had been becoming frequently discussed topics. More and more people pay attention to low-carbon lifestyle. According to experts predict 5% of China carbon emissions may originated from auto industry. However, China's auto industry is in a rapid development period, the proportion of carbon emissions will increase gradually. In automotive industry developed countries, 25%-28% of carbon emissions originated from auto industry.

In Copenhagen, China has promised to the world that carbon emissions per GDP will be decreased 40%-45% by 2020. It is the power and aim for all the vehicle

* Corresponding author.

manufacturers to develop and reform. In recent years, China's car sales volume had been growing rapidly [1]. At the same time, carbon emissions' influence on human living environment should not be ignored.

Meeting requirements of car sales forecast, and using Rough Set Theory's advantage that implicit knowledge can be found according to analysis and reasoning of data, this paper propose the process of China's car sales forecast based on Rough Set Theory; after reasonable choosing each statistical indicators, interpolating original data to steps of forecast process, the procedure forecasting China's car sales trend can be obtained; and then China's car sales trend was forecasted with this procedure; carbon emissions trend of China's vehicle segment was forecasted with auto industry carbon emissions calculation standard.

2 Introduction of Rough Set Theory

Rough Set Theory is a kind of powerful mathematical tools with which uncertain and incomplete information can be processed, its core idea is that an objective is described by some information, the objectives which are described by the same information cannot be distinguished, and this indiscernible relationship is the mathematical basic of Rough Set Theory [2].

3 Process of Car Sales Forecast Based on Rough Set Theory

Fig. 1 shows the process of car sales forecast based on rough set theory.

Step 1. Establishing system goal. Car sales growth rate over the years is selected as the system goal of China's car sales forecast, in other words, it is the decision attribute in China's car sales forecast decision tables.

Step 2. Setting up decision tables. Decision tables are the main study objects of rough set theory. They are mainly constituted by objects, attributes and their values. Objects and attributes (statistical indicators) should be determined based on system goal. Because of the influence of priori information, decision tables are set up uncertain.

Step 3. Initialization of pretreatment. First of all, divided the train set and test set, train set is used for getting rules; and test set is used for evaluating the applicability and accuracy of acquired rules; and then dimension is reduced to eliminating the redundant attributes added while decisions were set up and ensuring the proportion of the number of objects and attributes within a reasonable range.

Step 4. Pretreatment. It is divided to completion of decision tables and discretization of continuous attribute value. If there is the phenomenon of missing data in set up decision tables, it should be filled. When dealing with decision tables with Rough Set Theory, the attribute values must be the discrete data. Then continuous statistics are needed to be discretized. There are many ways to discrete continuous data [3][4], the results of those methods are different.

Step 5. Attributes and reduction of attributes values. It is good for the simplicity of obtained rules to reduce attributes values by a certain method. The number of obtained rules can be reduced by Reduction of attributes values, and then rules' applicability can be raised.

Step 6. Rules evaluation. Rules are evaluated by test set. If the accuracy and applicability of rules are not satisfactory, it is necessary to change the approach begin from one step of step 2 to step 5 and generate the new rule.

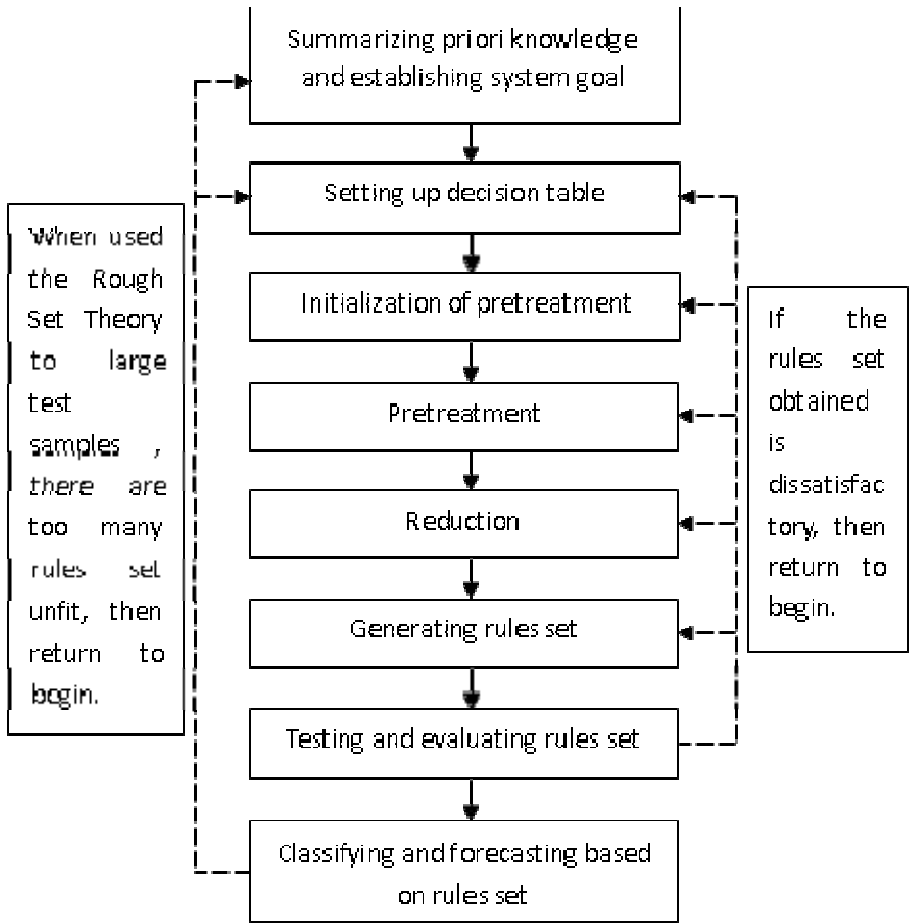


Fig. 1. The process of cars sales forecast based on Rough Set Theory

4 Car Sales Forecast Based on Rough Set Theory

4.1 Creation of China’s Cars Sales Index System

Two factors should be considered in the indicator selection process: One is whether statistical data corresponding to indicators can be obtained, it is up to the statistics habit in China; the other is indicators should be selected based on marketing, economic theory and statistical theory.

In this paper, the index system is created based on system goals and factors of population, economy, nature, technology, politics and laws.

4.2 Initialization of Pretreatment

In the process of China's cars sales forecast, the train set is selected the same as test set because of limited data amount. Some indicators are removed that duplication, intersect with others, non-quantifiable, have little effects to decision attribute indicators. And then the final index system is composed of proportion of urban population, GDP per capita, automobile consumption credit, Engel's Coefficient of urban households, disposable income per capita of urban households, consumer savings, the proportion of total energy production of oil, the proportion of total energy consumption of oil, mileage and cars sales.

In order to use Rough Set Theory to obtain the rule and grasp the dynamic relationship between each indicator, use the growth rate of each indicator above. (It is defined the growth rate of each indicator = [(the real value of each indicator in year $t+1$) - (the real value of each indicator in year t)] / (the real value of each indicator in year t) \times 100%). So the condition attributes of decision tables include the growth rate of proportion of urban population, GDP per capita, automobile consumption credit, Engel's Coefficient of urban households, disposable income per capita of urban households, consumer savings, the proportion of total energy production of oil, the proportion of total energy consumption of oil and mileage (They are represented by a, b, c, e, f, g, h, i and j.); decision attribute is the growth rate of cars sales (It is represented by d.).

Considered the system goal, availability of data and China's economic development process, each year from 1979-2009 as objects, it is 30 in all. Considered there may be some years (objects) with the same indicator growth rate (attribute value), they are combined into one object and its several times appearance should be described by a weight, and the weight is named frequency weight. In this paper, referred to journal article [5], the frequency weights of each object are set as 4.

Based on the growth rate of China's cars sales in each year from 1979, a partition is set that $(-\infty, 0]$, $(0, 10]$, $(10, 20]$, $(20, 30]$, $(30, +\infty)$: when the growth rate of cars sales at $(-\infty, 0]$, $d=0$; when the growth rate of cars sales at $(0, 10]$, $d=1$; when the growth rate of cars sales at $(10, 20]$, $d=2$; when the growth rate of cars sales at $(20, 30]$, $d=3$; when the growth rate of cars sales at $(30, +\infty)$, $d=4$.

4.3 Pretreatment

After data pretreatment by Rosetta software, the breakpoint set is obtained: $\{p_{10}^a, p_{16}^b, p_{25}^b, p_{10}^e, p_{16}^e, p_{15}^f, p_{11}^h, p_{10}^i\}$, $p_{10}^a=[0.74, 0.77]$, $p_{16}^b=[574, 687]$, $p_{25}^b=[1717, 1794]$, $p_{10}^e=[-0.9, -0.8]$, $p_{16}^e=[-0.2, 0.1]$, $p_{15}^f=[3446.26, 4710.55]$, $p_{11}^h=[-0.3, -0.2]$, $p_{10}^i=[-0.3, -0.2]$. And c(the growth rate of automobile consumption credit), g(the growth rate of disposable income per capita of urban households) and j(the growth rate of mileage) are removed because they are unrelated indicators. Table.1 is the decision table after data pretreatment.

Table 1. Decision table S_2

Year	No.	A	b	e	f	h	i	d	Frequency weight (number of times)
1979	1	1	0	1	0	1	0	3	1
1980	2	0	0	1	0	1	0	2	1
1981	3	1	0	1	0	0	0	0	1
1982	4	1	0	2	0	0	0	2	1
1983	5	0	0	2	0	0	0	3	1
1984	6	1	0	0	0	0	0	4	1
1985	7	0	0	0	0	1	0	4	1
1986	8	1	0	0	0	1	1	0	2
1987	9	1	0	2	0	1	1	3	2
1988	10	0	0	0	0	0	1	4	2
1989	11	0	0	2	0	0	1	0	2
1990	12	0	0	1	0	0	0	0	2
1991	13	0	2	1	0	1	1	4	4
1992	14	0	0	0	0	0	1	4	4
1993	15	0	1	0	0	1	1	3	4
1994	16	0	1	1	1	0	0	1	4
1995	17	0	1	2	1	0	1	1	4
1996	18	0	1	0	1	1	1	1	6
1997	19	1	0	0	1	1	1	1	6
1998	20	1	0	0	1	1	1	1	6
1999	21	1	0	0	1	0	1	2	6
2000	22	1	1	0	1	1	1	2	6
2001	23	1	1	0	1	0	0	2	8
2002	24	1	1	1	1	0	1	4	8
2003	25	1	1	1	1	0	0	4	8
2004	26	1	2	2	1	0	1	2	8
2005	27	1	1	0	1	0	0	2	8
2006	28	1	2	0	1	0	0	3	10
2007	29	1	2	2	1	0	0	3	10
2008	30	0	2	2	1	0	0	1	10

4.4 Getting the Rules

In this paper, the rule is obtained by Rosetta software, after comparing some methods, the method of Johnson’s algorithm is selected to obtain the rules. Table.2 is the obtained rules.

Table 2. Obtained rules

Conditions	Decisions
(1) $(b=h=0) \wedge e=1$; or (2) $(a=b=0) \wedge (e=2) \wedge (i=1)$; or (3) $(a=h=1) \wedge (e=f=0)$.	d=0
(4) $(a=0) \wedge (f=1)$; or (5) $(b=0) \wedge (f=h=1)$ or (6) $(b=1) \wedge (e=2)$.	d=1
(7) $(a=b=0) \wedge (e=h=1)$; or (8) $(a=1) \wedge (b=h=0) \wedge (e=2)$; or (9) $(a=b=1) \wedge (e=0)$; or (10) $(b=h=0) \wedge (f=1)$; or (11) $(a=i=1) \wedge (b=2)$.	d=2
(12) $(a=e=h=1)$; or (13) $(a=b=i=0) \wedge (e=2)$; or (14) $(e=2) \wedge (h=1)$; or (15) $(b=2) \wedge (e=0)$; or (16) $(b=1) \wedge (f=0)$; or (17) $(a=1) \wedge (b=2) \wedge (i=0)$.	d=3
(18) $(b=e=i=0)$; or (19) $(a=b=e=0)$; or (20) $(a=b=e=1)$; or (21) $(b=2) \wedge (e=1)$; or (22) $(e=i=1)$.	d=4

The evaluation of importance is made to the obtained rules. Table 3 is the calculation results of the rules set's importance.

Table 3. The calculation results of the rules set's importance

Rule No.	Definition	Applicability	Importance	Rule No.	Definition	Applicability	Importance
(1)	1.0	0.45	0.016	(12)	1.0	0.04	0.005
(2)	1.0	0.30	0.011	(13)	1.0	0.04	0.005
(3)	1.0	0.30	0.011	(14)	1.0	0.09	0.011
(4)	1.0	0.84	0.132	(15)	1.0	0.43	0.055
(5)	1.0	0.42	0.066	(16)	1.0	0.17	0.022
(6)	1.0	0.14	0.022	(17)	1.0	0.86	0.110
(7)	1.0	0.03	0.005	(18)	1.0	0.09	0.011
(8)	1.0	0.03	0.005	(19)	1.0	0.30	0.038
(9)	1.0	0.74	0.121	(20)	1.0	0.69	0.088
(10)	1.0	0.20	0.033	(21)	1.0	0.17	0.022
(11)	1.0	0.27	0.044	(22)	1.0	0.52	0.066

4.5 Rules Test

Based on "reuse method", the train set is used as test set. It is some optimism that doing like this, but it can meet the requirements of the rules test. Table 4 is rules test results of decision classification.

Table 4. Rules test results of decision classification

Decision values	Rate of correct rules classification	Rate of false rules classification	Indiscernible rate
0	100%	0	0
1	100%	0	0
2	71.4%	0	28.6%
3	50%	0	50%
4	85.7%	0	14.3%
Total	78.9%	0	21.1%

4.6 China's Car Sales Forecast in 2011

In 2000, China's urbanization rate is 36.2%, and the indicator achieved an average annual increase of 1 percentage points in the last 10 years. Reference to international experience, China's urbanization rate is expected to 49% and 56% in 2011 and 2020. China's average annual increase of urbanization rate is expected to 1.5 percentage points in the next 10 years, so the increase of urban population proportion is more than 0.74%, selected a=1.

During the Eleventh Five-Year, the increase rate of GDP is about 7.4%. According to statistics, China's GDP per capita is 3711USD in 2009 and more than 4000USD in 2010, and forecasted the GDP growth rate is in the range of 9%-9.5% in 2011. So it is sure that the increase of China's GDP per capita in 2011 is more than 1756RMB, selected b=2.

Engel's coefficient is the proportion of expense on food to the consumption expense. It is also said that the more you earn, the lower this number. In 2011, China's Engel's coefficient of urban/rural households is expected reduce to 30%/40%. So it is sure that the increase of Engel's coefficient of urban households is less than -0.7, selected $e=0$.

So far, the obtained condition attributes values are $a=1$, $b=2$, $e=0$. Finally, rule 3, 11, 15 or 17 in the rule set is selected.

After compared the evaluation results of rules importance indicator, rule 15 and 17 are considered more appropriately to forecast China's cars sales in 2011. So the 2011 increase rate of China's cars sales is forecasted in the interval (20,30], and 2010 China's cars sales is 18.0619 million, so it is forecasted that 2011 China's cars sales are in the range of 21.6743 to 23.4805 million.

5 2011 Add Vehicles Carbon Emissions Forecast

Cars carbon emissions are the product of fuel combustion. Based on the internationally recognized formula of carbon emissions: number of liters \times 2.7 (kg), cars carbon emissions increase can be calculated.

It is forecasted above that 2011 China's cars sales are in the range of 21.6743 to 23.4805 million, the one third of forecast interval as the most possible 2011 China's cars sales forecast value: 22.2764 million.

In the current, the lowest automobile fuel consumption per hundred kilometers is about 6L, and the highest is more than 20L, high fuel consumption cars are almost luxury cars and SUV, so the automobile consumption per hundred kilometers is selected as 10L in this paper. According to statistics, private cars' annual mileage is about 10-20 thousand kilometers; official cars' is about 20-40 thousand kilometers; commercial cars' (such as taxis) is about 100 thousand kilometers. In 2008, China's vehicle population of private cars, official cars and commercial cars are 35.0139 million, 15.9822 million and 9.3061 million. Referenced to this proportion, 2011 new cars' carbon emissions can be calculated.

In 2011, the add population of private cars will be 12.9337 million, all their annual mileage is 129.337-258.674 billion kilometers, its middle value 194.006 billion kilometers is be used to calculate; the add population of official cars will be 5.9033 million, all their annual mileage is 118.066-236.132 billion kilometers, its middle value 177.099 billion kilometers is be used to calculate; the add population of commercial cars will be 3.4394 million, all their annual mileage is about 343.940 billion kilometers.

After calculated, 2011 China's add vehicles carbon emissions are 193.06215 thousand ton.

6 Conclusions

The Rough Set Theory has emerged as a major mathematical method to manage uncertainties from inexact, noisy and incomplete information. It is an initial attempt of car sales forecast intelligentized that applied the Rough Set Theory to car sales

forecast. It has some references on developing policies of carbon emissions that forecasting added carbon emissions based on car sales forecast.

It should be noted that the rules set obtained in this paper can just apply for car sales forecast in recent years. Because there are some limits on the data acquisition, the data only before 2008 can be found to use in decision table. After finding the latest statistics, the more accurate forecast can be made based on the new rules set.

Finally, there are some suggestions on reducing carbon emissions: 1. Reforming the technologies on cars, upgrading the engine and transmission technologies to reduce carbon emissions. 2. Searching for new energy alternatives to fossil fuels actively. 3. Developing and promoting electric vehicles actively. 4. Developing some polices to limit the population of private cars. 5. Implementing carbon based road or purchase taxation systems for passenger cars. 6. Penalizing the high carbon emission vehicles with taxation.

References

1. Hu, W.: Research on Forecasting of Automobile Sales in China. Hohai University, Nanjing (2007)
2. Pawlak, Z.I.: Rough Set Approach to Knowledge-based decision support. *European Journal of Operational Research* 99(1), 48–57 (1997)
3. Nguyen, S.H., Skowron, A.: Quantization of Real Value Attributes-Rough Set and Boolean Reasoning Approach. In: Proc. of the Second Joint Annual Conference on Information Sciences, Wrightsville Beach, North Carolina (1995)
4. Tian, F., Zhang, H., Lu, Y., Shi, C.: Bayes Classification Approaches with Continuous Attributes. *J. Tsinghua Univ (Sci. & Tech.)* 43(1), 75–78 (2003)
5. Tang, J.: Reliability Analysis to Deal with Imperfect Information by Rough Set Theory. *J. Control and Decision* 17(2), 255–256 (2002)

The Application on Attribute Reduction by Using Bacterial Foraging Optimization and PSO Algorithm

Wang Jianguo*

Department of Computer Science and Technology, Xinzhou Teachers University,
Xinzhou 034000, Shanxi, China
chinawjg@yahoo.cn

Abstract. Inspired by the bacteria foraging process, the thought of bacteria foraging algorithm used in particle swarm algorithm, this paper puts forward a kind of bacteria foraging particle swarm algorithm. For the trend of the operation process of bacteria can guide particles toward the more optimal direction evolve, and the particle swarm algorithm and improve bacteria foraging algorithm convergence speed and optimization ability. And the algorithm is applied to attribute reduction. Numerical results show that the proposed bacteria foraging particle swarm optimization algorithm of the reduction in optimization ability, are better than Hu algorithm, particle swarm reduction algorithm and bacteria foraging reduction algorithm, can get a better minimum attribute reduction.

Keywords: Attribute reduction, bacterial foraging optimization algorithm, PSO.

1 Introduction

How to establish a more self-contained knowledge database fast and efficiently is considered publicly as a bottleneck problem in an expert system . There may be a few of redundant and no relative attributes in the collected cases data.

In 1982, Z.Pawlak published the classic paper rough sets, which indicated the birth of rough set theory. Z.Pawlak's monograph published in 1991 and application monograph published in 1992 were better summarized of the theory and practical achievement in this period and promoted the application of rough set in every fields[1]. By 20 years development there has been plentiful and substantial achievement either in rough set theory or in application. Every year there are international conferences held and many articles delivered about rough set theory, and also several books have been published.

Attribute reduction which is one of research fields in rough set, has long been an important research direction for data mining. Under the condition of keeping classification system or decision-making ability invariable, not important and redundant attributes deleted from the decision-making system to reduce the volume of data to handle, and to improve the simplicity of data mining. Fewer attributes can be used to express the original properties of the same classification or decision-

* Corresponding author.

making ability. However, obtaining the minimum reduction in decision table has been proved to be NP problem [2,3]. In the process of researching attribute reduction, people usually deal with those decision-making tables which an attribute value is completely accurate. However, due to various objective factors, the reality data exists in the impact of missing, which is often difficult to ensure that the collected data are complete and accurate. Thus, attribute reduction to incomplete information system is more practical significance.

In recent years, some scholars have researched in attribute reduction for incomplete information system, such as HU, etc [4]. They proposed the attribute reduction methods for incomplete information system based on the Conditional Information Entropy in attribute importance. Some scholars apply intelligence Optimization in incomplete information systems attribute reduction. They proposed methods which genetic algorithm and particle swarm optimization (PSO) combined with the attribute reduction [5-9]. In PSO, particles easily trapped in local optimal point, the entire population of global optimal search in the vicinity, and the search area is limited, which is easy to fall into local optimum, and difficult to get a good minimum attribute reduction.

In response to these limitations, this paper proposed a hybrid intelligent algorithm which combines bacterial foraging that a bionic random search algorithm put forward by K. M. Passino with particle swarm optimization algorithm and applied to attribute reduction for incomplete decision table. By using Hu algorithm, particle swarm reduction algorithm, bacterial foraging reduction algorithm and bacteria foraging optimization and particle swarm reduction algorithm (BFPSO) on the UCI, experimental results show that the bacteria foraging particle swarm optimization reduction algorithm has better ability in the solution space and has a strong ability to jump out of local extreme point, and maintain the fast convergence rate. The new algorithm easily obtains better the minimum relative attributes reduction.

2 Bacterial Foraging Optimization Algorithm

In 2001, Prof. K. M. Passino proposed an optimization technique known as Bacterial Foraging Optimization Algorithm (BFOA) based on the foraging strategies of the E.Coli bacterium cells [15]. Until date there have been a few successful applications of the said algorithm in optimal control engineering, harmonic estimation [16,21], transmission loss reduction [17], machine learning [18-20] and so on.

Bacterial Foraging Optimization Algorithm includes three kinds of operations, such as chemotaxis, reproduction and migration (elimination dispersal). In the search space, BFO algorithm seeks the best solution through bacteria chemotaxis operation and realizes quorum sensing by the collection function of bacteria. Using the realization of "survival of the fittest" reproduction operation this evolution rules, and Use the eliminate mechanism to avoid into maturity.

For an E-coli bacterium, Chemotaxis operation can simulate by two different ways alternatively: tumble and run. In bacteria during the whole life has been accompanied by these two kinds of operation. A tumble is represented by a unit walk with random direction, a unit walk with the same direction as the previous step indicates a run. it is two kinds of alternate ways to convert the bacteria, which find moves nutrients, namely, the optimal solution.

In the evolutionary process, elimination and dispersal events can occur such that bacteria in a region are killed or a group is dispersed into a new part of the environment due to some influence. They have the effect of possibly destroying chemotactic progress, but they also have the effect of assisting in chemotaxis, since dispersal may place bacteria near good food sources. From the evolutionary point of view, elimination and dispersal was used to guarantee diversity of individuals and to strengthen the ability of global optimization. In BFOA, bacteria are eliminated with a probability of ped . In order to keeping the number of bacteria in the population constant, if a bacterium is eliminated, simply disperse one to a random location on the optimization domain.

3 Particle Swarm Optimization Algorithm

Particle swarm optimization algorithm [12](PSO) is a kind of biological evolution algorithm, which simulate the predator behaviour of bird to achieve the optimization solution of optimization problem. First, in the solution space randomly initialised birds group. In bird group's each bird is called "particle". After undergoing certain iterations finds the optimal solution. All the particles are in search of the solution space, each with its own particle position and speed, expressing with vectors.

Because of PSO algorithm is a new kind of swarm intelligence algorithm, and has been proved to be an effective global optimization method. The proposed algorithm is applied PSO for minimum attribute reduction [12, 13], and has obtained good results.

Comparing with other evolutionary algorithms such as Genetic Algorithm, PSO converge faster. The convergence efficiency of PSO is influenced little by the number of dimensions. However, huge number of dimensions may cause some other problems, such as local optimization. It means particles prematurely converge at local optimal position and stop searching for the real globally optimum.

However, the different of kinds of attribute reduction algorithm based on PSO is mainly embodied in representation methods of problem and the structures of the fitness function. Commonly used treatment method is that problem was transformed into single-object problem to solve using the penalty function. But in practice, how to choose a suitable punishment factor will be far from an easy proposition, but will greatly affect yield of optimal solution, so get not least attribute reduction. Besides PSO algorithm is easy to jump into and local optimal to jump. Using bacteria foraging optimization algorithm combining with the PSO algorithm applied to the attribute reduction, thus to avoid being trapped in local optimum and get the optimal attribute reduction.

4 BFPSO Attribute Reduction Algorithm

Based on the optimization algorithm and bacteria foraging particle swarm algorithm, this paper presents a method to combine both the hybrid intelligence algorithm applied to the Attribute Reduction, namely bacteria foraging optimization and particle swarm algorithm combines the Attribute Reduction (name as BF - PSO Attribute

Reduction), the two methods can compensate deficiencies through the tendency of bacteria foraging algorithm realization mobile operators, while local search using PSO operation within the search space of the global search, and get better Attribute Reduction.

4.1 BF-PSO Algorithm

The BF-PSO combines both algorithms BF and PSO. The aim is to make use of PSO ability to exchange social information and BF ability in searching a new solution by elimination and dispersal. In BFOA, a unit length direction of tumble behaviour is randomly generated. Random direction may lead to delay in reaching the global solution. In the BF-PSO, the unit length random direction of tumble behavior can be decided by the global best position and the best position of each bacteria. During the chemotaxis loop, the update of the tumble direction is determined by:

$$\phi(j+1) = w * \phi(j) + C_1 * R_1(pbest - pcurrent) + C_2 * R_2(gbest - pcurrent) \tag{1}$$

Where Pbest is the best position of each bacterial and gbest is the global best bacterial. This algorithm in BF - PSO can reduce randomness, and improve group performance.

4.2 Initialization Code

Application of bacterial foraging and particle swarm optimization (PSO) combining to solve the incomplete decision table attribute reduction must firstly solve coding problem. Designs a suitable expression solution the method, whether in early or in the evolutionary process of all individuals will be feasible, will be the influence entire algorithm each stage committed step.

According to incomplete decision attribute table, we can make the minimum attribute reduction problem equivalently transformed to the 0-1 combinatorial optimization problem:

$$\min C(x) = x_1 + x_2 + \dots + x_N \tag{2}$$

Where $x \in \{0,1\}^m, \alpha(c) = \alpha(x), \gamma(c) = \gamma(x)$.

In the application of BF - PSO algorithm for solving optimization problem, a Boolean vector represents bacteria. Various bacteria corresponding condition attribute set $C = (c_1, c_2, \dots, c_N)$ of incomplete decision-making table, bacterium indicated with the binary string which length is m . Each of the corresponding decision table is a condition attribute of a binary string, when the value is 1, said choose corresponding condition attribute, conversely, is said to be the condition attribute reduction. For example: In the decision table includes six condition attributes $(c_1, c_2, c_3, c_4, c_5, c_6)$, but it relatively smallest reduction is (c_1, c_4, c_5) , delete the reduction (c_2, c_3, c_6) , then the bacterium individual may express 100110.

4.3 The Fitness Function

Fitness function directly determines the searching behaviour of bacterial populations, and the fitness function is the only certainty indicators to evaluate the adaptability of bacteria. So selecting the appropriate fitness function is particularly important.

Classification quality and classification accuracy plays a very important role in the attribute reduction, the higher classification quality and classification accuracy, the better the classification. Therefore, the fitness function is defined as follows:

$$F(x) = k_1(1 - \frac{C(x)}{|U|}) (k_2 \frac{\sum_{i=1}^N |Y_i c|}{\sum_{i=1}^N |Y_i|} + k_3 \frac{\sum_{i=1}^N |Y_i c|}{|U|}) \tag{3}$$

where $\frac{\sum_{i=1}^N |Y_i c|}{\sum_{i=1}^N |Y_i|}$ is approximate the classification accuracy, $\frac{\sum_{i=1}^N |Y_i c|}{|U|}$ is approximate classification accuracy, $C(x)$ is the number of condition attributes after reduction.

In addition, because each part of the importance of the attribute reduction is different, the front by adjusting the parameters k_1, k_2 of the function can make the fitness function more efficient to accurately. The sizes of the fitness of bacteria decide by the classification quality and classification of the level of decision accuracy. The higher the classification accuracy or classification quality, the greater the fitness value is.

4.4 BF-PSO Attribute Reduction Algorithm

Using BF - PSO for incomplete information of the decision table attribute reduction algorithm procedure is as follows:

Input: incomplete information decision table $S=(U, C \cup D, V, f)$, the size of Bacterial N , the number of chemotaxis N_c , the number of reproduction N_{re} , the number of elimination dispersal N_{ed} , mobile steps N_s , the eliminate probability P_{ed} , PSO parameters C_1, C_2, R_1, R_2, w .

Output: Optimal attribute reduction.

Step1: Calculation of the Core attributes Core(C). Core attributes don't participate in the solution space search because Core attributes is non-missing attributes in the attribute reduction. This can reduce the solution space dimension to improve the efficiency of reduction.

Step2: Calculating the weight of each attribute.

Step3: According to the weights of attribute, initialization population and parameter such as the dimension of search space p , the number of bacterial species N , the number of chemotaxis N_c , the number of reproduction N_{re} , the number of elimination dispersal N_{ed} , mobile steps N_s , the eliminate probability P_{ed} , the

step $C(i), i=1,2,\dots,N$ in forward direction $\phi(j)$, PSO parameters C_1, C_2, R_1, R_2, w .

Step4: Calculate individual fitness values, the initial pbest of each bacteria coordinate set has been initialized value of the current weight of bacteria, compared the fitness value of all bacteria, record number of bacteria the optimal value, the optimal location of bacteria is set to gbest.

Step5: Elimination dispersal loop: $l=l+1$.

Step6: Reproduction loop: $k=k+1$.

Step7: Chemotaxis loop: $j=j+1$, first each bacterial implement one step of chemotaxis, and then calculate the corresponding fitness function $J(i, j, k, l)$, let $J_{last}=J(i, j, k, l)$,

$$\phi(j+1) = w * \phi(j) + C_1 * R_1(pbest - pcurrent) + C_2 * R_2(gbest - pcurrent) \quad \bullet$$

$$\theta^i(j+1, k, l) = \theta^i(j, k, l) + C(i)\phi(j) \text{ Computing fitness function } J(i, j, k, l).$$

Step8: If $j < N_c$, goto step7.

Step9: Reproduction: calculate health function for each bacterial $J_{health} = \sum_{j=1}^{N_c+1} J(i, j, k, l)$, (higher cost means lower

health)•Sort bacteria and chemotactic parameters $C(i)$ in order of ascending cost. For keep a constant population size, bacteria with the highest J_{health} values die. The remaining bacteria are allowed to split into two bacteria in the same place to keep.

Step10: If $k < N_{re}$, go to step6.

Step11: Elimination-dispersal: Eliminate and disperse bacteria with probability P_{ed} ,

Step12: If $l < N_{ed}$, then go to step5, else end algorithm.

Step13: gbest is the optimal attribute reduction.

5 Experimental Result Analysis

In order to investigate the validity and correctness of the algorithm, this paper tests five standard data sets given in UCI machine learning repository [14]. Experimental environments are:

(1) CPU: E8500 3.16GHz. (2) Memory: 4GB. (3) Windows XP, Matlab7.0.

Computing respectively Hu algorithm, particle swarm reduction algorithm and bacterial foraging swarm reduction algorithm on the UCI data, the experimental results of those algorithms shown in Table 1.

Table 1 shows that BF-PSO algorithm is better than Hu algorithm, PSO and BFO reduction algorithm in the reduction results. BF-PSO reduction algorithm can obtain less attributes reduction. This is because the BF-PSO algorithm uses the characteristics of bacterial foraging. In foraging process, the bacteria accept continual pheromone stimulation, so we have more of the search space and faster convergence.

Table 1. Reduction results

data set	number of attribute	Record number	algorithm	Number of attribute reduction	Running time /s
Sponge	45	76	Hu	8	1.647
			PSO	12	6.294
			BFO	11	11.765
			BF-PSO	8	10.987
Zoo	17	101	Hu	6	1.020
			PSO	6	1.670
			BFO	6	2.783
			BF-PSO	5	2.670
Lym phography	19	148	Hu	6	2.017
			PSO	7	21.389
			BFO	7	32.786
			BF-PSO	6	31.561
Soybean large	36	307	Hu	10	5.978
			PSO	12	46.742
			BFO	11	100.97
			BF-PSO	10	96.713
Vote	17	435	Hu	9	3.470
			PSO	10	13.276
			BFO	10	19.721
			BF-PSO	8	18.642

BF-PSO combined with the direction of bacteria foraging constantly adjust the movement direction of the bacteria. If the new individual we obtained is better than the current worst individual in population, then the current worst individual replaced with the current individual for making the average fitness degree of entire group increase. In addition, the outstanding individual has been retained in the group, and the new algorithm increases the fine genes of group. All the reasons make it find optimal or satisfactory solution. Therefore, BF-PSO algorithm has better in optimization.

As for running time of these algorithms, Hu algorithm use less time, and $Hu < PSO < BF-PSO < BFO$. The reason is that the Hu algorithm adopts the attribute weight priority selection greedy strategy; it can very quickly obtain the reduction result. But the Hu algorithm obtains the result frequently is not the smallest attribute reduction. In addition, the BF-PSO algorithm's running time lay in the BF-PSO reduction algorithm compared to the PSO algorithm long reason is introduces the migration operation in the PSO algorithm foundation, the bacterium unceasing migration has caused the BF-PSO reduction algorithm to consume time long somewhat, but the BF-PSO algorithm expanded the search space. Because we mentioned four algorithms can obtain in such a short time most give preferential treatment to service families the reduction, but the BF-PSO reduction algorithm has the better optimization ability compared to other reduction algorithm, can obtain the least attribute reduction.

References

1. Zhang, W.-x., Wu, W.-z., Liang, J.-y.: *Rough Set Theory and Methods*. Science Press, Beijing (2001)
2. Wu, M.-f., Xu, Y., Liu, Z.-m.: Heuristic Algorithm for Reduction Based on the Significance of Attributes. *Journal of Chinese Computer Systems* 28(8), 1452–1455 (2007)
3. Chen, Y., Xu, X.-h., et al.: Study of Modified Particle Swarm Optimization Algorithm Based on Immune Clone Principle. *Journal of System Simulation* 20(6), 1471–1774 (2008)
4. Hu, X.-h., Cercone, N.: Learning in Relational Database: A Rough Set Approach. *Computational Intelligence* 11(2), 323–337 (1995)
5. Ren, Y.-g., Wang, Y., Yan, D.-q.: Rough Set Attribute Reduction Algorithm Based on GA. *Journal of Chinese Computer Systems* 27(5), 862–865 (2006)
6. Liao, J.-k., Ye, D.-y.: Minimal attribute reduction algorithm based on particle swarm optimization with immunity. *Journal of Computer Applications* 27(3), 550–555 (2007)
7. Wang, X.-y., Yang, J., Teng, X.-l.: Feature selection based on rough sets and particle swarm optimization. *Pattern Recognition Letters* 28, 459–471 (2007)
8. Yang, X.-m., Yuan, J.-s., Yuan, J.-y.: 'A modified particle swarm optimizer with dynamic adaptation. *Applied Mathematics and Computation*' 189, 1205–1213 (2007)
9. Guo, J.-l., Wu, Z.-j., Jiang, D.-z.: Adaptive Swarm Optimization Algorithm Based on Energy of Particle. *Journal of System Simulation* 21(5), 4465–4471 (2009)
10. Olesen, J.R., Cordero H., J., Zeng, Y.: Auto-clustering using particle swarm optimization and bacterial foraging. In: Cao, L., Gorodetsky, V., Liu, J., Weiss, G., Yu, P.S. (eds.) *ADMI 2009. LNCS*, vol. 5680, pp. 69–83. Springer, Heidelberg (2009)
11. Chen, H.N.: Cooperative Bacterial Foraging Algorithm for Global Optimization. In: *Proceedings of CCDC 2009: 21st Chinese Control and Decision Conference*, vol. 1-6, pp. 3896–3901. IEEE, New York (2009)
12. Chapman, B., Hernandez, O., Huang, L., Weng, T.H., Liu, Z., Adhianto, L., Wen, Y.: 'Dragon: an open64-based interactive program analysis tool for large applications. To appear in the *Proceedings of the 4th International Conference on Parallel and Distributed Computing, Applications and Technologies, PDCAT* (2003)
13. Chapman, B., Mehrotra, P., Zima, H.: Enhancing openMP with features for locality control. In: *Proc. ECMWF Workshop Towards Teracomputing – The Use of Parallel Processors in Meteorology*, Reading, England (November 1998)
14. UCI. Repository of machine learning databases [DB/OL] (July 8, 2007), <http://www.ics.uci.edu/~mllearn/MLRepository.html>
15. Passino, K.M.: Biomimicry of Bacterial Foraging for Distributed Optimization and Control. *IEEE Control Systems Magazine*, 52–67 (2002)
16. Mishra, S.: A hybrid least square-fuzzy bacterial foraging strategy for harmonic estimation. *IEEE Trans. on Evolutionary Computation* 9(1), 61–73 (2005)
17. Tripathy, M., Mishra, S., Lai, L.L., Zhang, Q.P.: Transmission Loss Reduction Based on FACTS and Bacteria Foraging Algorithm. In: *PPSN*, pp. 222–231 (2006)
18. Kim, D.H., Cho, C.H.: Bacterial Foraging Based Neural Network Fuzzy Learning. In: *IICAI*, pp. 2030–2036 (2005)
19. Holland, J.H.: *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Harbor (1975)
20. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: *Proceedings of IEEE International Conference on Neural Networks*, pp. 1942–1948 (1995)
21. Kim, D.H., Abraham, A., Cho, J.H.: A hybrid genetic algorithm and bacterial foraging approach for global optimization. *Information Sciences* 177(18), 3918–3937 (2007)

A Novel Feature Selection Method for the Conditional Information Entropy Model

Jing Ruan¹ and Changsheng Zhang^{2,*}

¹ Wenzhou Vocational and Technical College, Wenzhou, 325035, P.R. China

² College of Physics and Electronic Information Engineering, Wenzhou University,
Wenzhou, 325035, P.R. China

{jsj_zcs, ruanjing79}@126.com

Abstract. In this paper, a novel feature selection method of discernibility object pair set is provided. At first, the feature selection definition of new method is presented. What's more, it is proved that the above feature selection definition is equal to the feature selection definition based on conditional information entropy. In order to compute discernibility object pair set, a quick algorithm for simplified decision system is introduced, whose time complexity is $O(C \| U)$. On this condition, an efficient and novel algorithm based on discernibility object pair set for feature selection in conditional information entropy model is designed, whose time and space complexity are $O(C \| U) + O(C \| U / C^2)$ and $O(U / C^2) + O(U)$ respectively. At last, an example is employed to illustrate the efficiency of the new algorithm.

Keywords: Rough set, Information entropy, Discernibility object, Pair set, Feature selection, Algorithm complexity.

1 Introduction

Rough set theory[1-3] is a valid mathematical tool to deal with vague and uncertain knowledge. It has been applied successfully in many practical fields, such as artificial intelligence, pattern recognition, data mining and intelligent decision-making area. Feature selection is one of the most important researches in rough set theory. Some researchers have proposed different feature selection algorithms according to different requirements. At present, some scholars have put forward the feature selection based on the information entropy model from the perspective of information theory, using information entropy as heuristic information to design feature selection algorithms. For example, a feature selection algorithm based on the conditional information entropy with the time complexity $O(C \| U^2) + O(U^3)$ in paper[4] is provided. While in paper[5] the algorithm with the time complexity $O(C^2 | U | \log | U |)$ is proposed as well. In recent years, some scholars use the methods of discernibility matrix[6,7] and binary discernibility matrix[8] to design feature selection algorithms. However, the disadvantage of the discernibility matrix and binary discernibility matrix is that it is

* Corresponding author.

first to construct the discernibility matrix or the binary discernibility matrix, for large decision system, this process is not only time-consuming, but also needs large storage space, so that the efficiency of the algorithm is reduced greatly.

In order to reduce the storage space of the discernibility matrix as possible as one can, and to use the designing idea of the discernibility matrix, in this paper, we combine the method of discernibility object pair set[9], a feature selection algorithm based on conditional information entropy model is designed. The proposed algorithm does not to construct the discernibility matrix, but also making use of the idea of the discernibility matrix. To cut down the time complexity of the new algorithm, based on the simplified decision system, firstly, a feature selection definition based on discernibility object pair set is presented, and it is proved that this definition is equal to the feature selection definition based on conditional information entropy. On this condition, algorithm based on discernibility object pair set in the model is designed, whose time and space complexity of the new algorithm are $O(|C||U|)+O(|C||U/C|^2)$ and $O(|U/C|^2)+O(|U|)$ respectively.

2 Preliminaries

In this section, we review some basic concepts and corresponding definitions, and introduce the notion of discernibility object pair set in complete decision systems.

Definition 2.1. A decision system is defined as $S = (U, C, D, V, f)$, where

- (1) $U = \{x_1, x_2, \dots, x_n\}$ is a non-empty finite set of objects;
- (2) $C = \{c_1, c_2, \dots, c_r\}$ is the set of condition features, D is the set of decision features, $C \cap D = \emptyset$;
- (3) $V = \bigcup_{a \in C \cup D} V_a$, where V_a is the value range of feature a ;
- (4) $f: U \times C \cup D \rightarrow V$ is an information function, which is an information value for each feature of each object, that is $\forall a \in C \cup D, x \in U, f(x, a) \in V_a$ holds.

With any feature subset $B \subseteq (C \cup D)$ there is a binary indiscernibility relation $IND(B) : IND(B) = \{(x, y) \in U \times U \mid \forall a \in B, f(x, a) = f(y, a)\}$, the partition of U , generated by $IND(B)$ is denoted as $U / IND(B)$ (for short U / B). Any element $[x]_B = \{y \mid \forall a \in B, f(x, a) = f(y, a)\}$ in U / B is called equivalence class.

Definition 2.2. Given a decision system $S = (U, C, D, V, f)$, any feature set $B \subseteq C \cup D$ (knowledge, equivalent relation, $U / B = \{B_1, B_2, \dots, B_t\}$) in U is a random variable in algebra formed by subsets of U , the probability distribution of the random variable can be determined by:

$$[B : p] = \begin{bmatrix} B_1 & B_2 & \dots & B_t \\ p(B_1) & p(B_2) & & p(B_t) \end{bmatrix}$$

Where $p(B_i) = |B_i| / |U|, i = 1, 2, \dots, t$.

Definition 2.3. [4] Given a decision system $S = (U, C, D, V, f)$, the conditional entropy of set of decision features D ($U / D = \{D_1, D_2, \dots, D_k\}$) to that of conditional features C ($U / C = \{C_1, C_2, \dots, C_m\}$) is defined as:

$$H(D | C) = - \sum_{i=1}^m p(C_i) \sum_{j=1}^k p(D_j | C_i) \log p(D_j | C_i)$$

Where $p(D_j | C_i) = |D_j \cap C_i| / |C_i|$.

Definition 2.4. [4] Given a decision system $S = (U, C, D, V, f)$ and $\forall b \in B \subseteq C$, if $H(D | B) = H(D | (B - \{b\}))$, then b is unnecessary for D in B ; else b is necessary for D in B . For $\forall B \subseteq C$, if every element in B is necessary for D , then B is independent to D .

Definition 2.5. [4] Given a decision system $S = (U, C, D, V, f)$, if $\forall B \subseteq C$, $H(D | B) = H(D | C)$, and B is independent to D , then B is called the feature selection of C to D (the feature selection based on information entropy).

Definition 2.6. Given a decision system $S = (U, C, D, V, f)$, if $\forall B \subseteq C, \forall x \in U$, $U / D = \{D_1, D_2, \dots, D_k\}, \forall x \in U$, we define the probability distribution function as follows:

$$\mu_B(x) = \left(\frac{|D_1 \cap [x]_B|}{|[x]_B|}, \frac{|D_2 \cap [x]_B|}{|[x]_B|}, \dots, \frac{|D_k \cap [x]_B|}{|[x]_B|} \right).$$

Definition 2.7. [9] Given a decision system $S = (U, C, D, V, f)$, denote $U / C = \{[x'_1]_C, [x'_2]_C, \dots, [x'_m]_C\}$ and $U' = \{x'_1, x'_2, \dots, x'_m\}$, then decision system $S' = (U', C, D, V, f)$ is the simplified decision system.

Definition 2.8. Given a decision system $S = (U, C, D, V, f)$, $S' = (U', C, D, V, f)$ is the corresponding simplified decision system, for $\forall P \subseteq C$, he discernibility object pair set is defined as follows: $R_p = \{ \langle x_i, x_j \rangle | x_i, x_j \in U', \exists a \in P, f(x_i, a) = f(x_j, a) \wedge \mu_p(x_i) \neq \mu_p(x_j) \}$, where $\langle x_i, x_j \rangle$ is called the discernibility object pair, R_p is the discernibility object pair set obtained by condition feature set P .

Definition 2.9. Given a decision system $S = (U, C, D, V, f)$, $S' = (U', C, D, V, f)$ is the corresponding simplified decision system S , for $\forall P \subseteq C$, if $R_p = R_C$ and $\forall a \in P \Rightarrow R_{P-(a)} \neq R_C$, then is called the feature selection of C with respect to D (the feature selection based on discernibility object pair set).

3 The Equivalence of Two Feature Selection

Theorem 1^[10]. Given a decision system $S = (U, C, D, V, f)$, for $\forall P \subseteq C$, denote $U / P = \{P_1, P_2, \dots, P_l\}$, $U / D = \{D_1, D_2, \dots, D_k\}$, then the sufficient and necessary condition of $H(D | P) = H(D | C)$ is that $\forall x \in U \Leftrightarrow \mu_p(x) = \mu_c(x)$.

Theorem 2. Given a decision system $S = (U, C, D, V, f)$, for $\forall P \subseteq C$, if $R_p = R_c$, there is $H(D | P) = H(D | C)$.

Proof: Suppose the result is not true, namely there is $H(D | P) \neq H(D | C)$. According to Theorem 1, we know $\exists x'_i \in U \Rightarrow \mu_p(x'_i) \neq \mu_c(x'_i)$. So there is $[x'_i]_p \neq [x'_i]_c$. Suppose $[x'_i]_p = [x^0]_c \cup [x^1]_c \cup [x^2]_c \dots \cup [x^l]_c$, (where $[x^i]_c \cap [x^j]_c = \emptyset$ ($i, j = 1, 2 \dots l$), $x'_i = x^0, l \geq 1$), so there exist $x_j \in \{x^1, x^2, \dots, x^l\}$ such that $\mu_p(x_j) \neq \mu_c(x'_i)$. Otherwise, there is $\mu_c(x'_i) = \mu_c(x^1) = \mu_c(x^2) = \dots = \mu_c(x^l)$, according to the proof process of Theorem 1, there is $\mu_p(x_j) = \mu_c(x'_i)$, which conflicts with $\mu_p(x_j) \neq \mu_c(x'_i)$; On the other hand, due to $[x_j]_c \cap [x'_i]_c = \emptyset$, so there exist $\exists c_h \in P \subseteq C$, such that $f(x'_i, c_h) = f(x_j, c_h) \wedge \mu_c(x'_i) \neq \mu_c(x_j)$, according to Definition 2.8, there is $\langle x'_i, x_j \rangle \in R_c$ holds, while $\mu_p(x'_i) = \mu_p(x_j)$, thus $\langle x'_i, x_j \rangle \notin R_p$, so there is $R_p \neq R_c$, which conflicts with the hypothesis $R_p = R_c$. So the hypothesis does not hold. This completes the proof.

Theorem 3. Given a decision system $S = (U, C, D, V, f)$, $S' = (U', C, D, V, f)$ is the simplified decision system of decision system S , for $\forall P \subseteq C$, if $H(D | P) = H(D | C)$, there is $R_p = R_c$.

Proof: Suppose the result is not true, namely there is $R_p \neq R_c$, there exist $\langle x'_i, x_j \rangle \in R_c$ and $\langle x'_i, x_j \rangle \notin R_p$. According to Definition 2.8, there exists feature $\exists c_h \in P \subseteq C$ and $x'_i \in [x'_i]_c \wedge x'_j \in [x'_j]_c$, such that $f(x'_i, c_h) = f(x'_j, c_h) \wedge \mu_c(x'_i) \neq \mu_c(x'_j)$, so there are $[x'_i]_c \neq [x'_j]_c$ and $[x'_i]_c \cup [x'_j]_c \subseteq [x'_i]_p$; On the other hand, due to $H(D | P) = H(D | C)$, according to Theorem 1, there is $\mu_p(x'_i) = \mu_c(x'_i) \wedge \mu_p(x'_j) = \mu_c(x'_j)$ holds; Due to $[x'_i]_c \cup [x'_j]_c \subseteq [x'_i]_p$. According to Definition 2.6 (the definition of the

probability distribution function)there is $\mu_p(x'_i) = \mu_p(x'_j)$, thus $\mu_c(x'_i) = \mu_c(x'_j)$, which conflicts with the hypothesis $\mu_c(x'_i) \neq \mu_c(x'_j)$, So the hypothesis does not hold, thus $R_p = R_c$ holds. This completes the proof.

Theorem 4. The definition of feature selection based on information entropy is equal to the definition of feature selection based on discernibility object pair set.

Proof: The proof follows directly from Theorem 2 and Theorem 3.

Theorem 4 states that the definition of feature selection based on information entropy is equal to the definition of feature selection based on discernibility object pair set. Therefore, to compute the feature selection based on information entropy can be built on the discernibility object pair set. This theorem provides a theoretical basis for us to design a new feature selection algorithm.

4 Feature Selection Method for Conditional Information Entropy Model

To calculate the simplified decision system is in fact to calculate the indistinguishable relation $IND(C)$. We use radix sorting to design a quick algorithm for computing the simplified decision system, whose time complexity is $O(|C| \parallel U |)$ [10].

4.1 Feature Selection Method Based on Conditional Information Entropy Model

Input: simplified decision system $S = (U', C, D, V, f)$, $C = \{c_1, c_2, \dots, c_s\}$;

Output: feature selection R ;
Begin

Step1: according to the algorithm of paper[10] to compute: $U' = (x'_1, x'_2, \dots, x'_m)$ and
for $\forall x \in U', \mu_c(x)$;

Step2: Sort $M_i (1 \leq i \leq s)$ from big to small by quick sorting, then
obtain: $M_{i1} \geq M_{i2} \geq \dots \geq M_{is}$;

the corresponding features are $c_{i1}, c_{i2}, \dots, c_{is}$, let $R = \{c_{i1}\}$; $R_c = \emptyset$;

Step3: To compute $U' / R = \{A_1, A_2, \dots, A_L\}$ by radix sorting;

Step4: For each $A_j = \{x_{j1}, x_{j2}, \dots, x_{jr}\} \in U' / R, (1 \leq j \leq L)$

for ($h = 1; h < r; h++$)

for ($g = h + 1; g < r + 1; g++$)

if ($\mu_c(x_{jh}) \neq \mu_c(x_{jg})$)

then incorporate $\langle x_{jh}, x_{jg} \rangle$ in R_c ;

```

Step5: for( $k = 2; k < s + 1; k ++$ )
    { if( $R_C = \emptyset$ ) the algorithm terminates;
      else  $R = R \cup \{c_{ik}\}$ ;
        For any discernibility object pair  $\langle x_h, x_t \rangle$  in  $R_C$ ;
          if( $f(x_h, c_{ik}) \neq f(x_t, c_{ik})$ )
             $R_C = R_C - \{\langle x_h, x_t \rangle\}$ ;
        }
End
    
```

4.2 Time and Space Complexity Analysis of the Method

In this section, the time and space complexities of the proposed algorithms are analyzed. The time and space complexities of each step in the feature selection algorithm are shown in Table 1.

Table 1. Time and space complexities of the feature selection algorithm

Step No.	Time complexity	Space complexity
Step 1	$O(C U)$	$O(U)$
Step 2	$O(C \log C)$	$O(\log C)$
Step 3	$O(C U' / C)$	$O(U')$
Step 4	$O(A_1 ^2 + A_2 ^2 + \dots + A_L ^2) \leq O(A_1 + A_2 + \dots + A_L)$ $= O(U' / C)^2$	$O(U' / C)^2$
Step 5	$O(C (A_1 ^2 + A_2 ^2 + \dots + A_L ^2))$, Loop $ C - 1$ times. The result is $O(C R_{C-\{c_{i1}\}}) \leq O(C (A_1 ^2 + A_2 ^2 + \dots + A_L ^2))$	$O(R_{C-\{c_{i1}\}})$
Total	$O(C U) + O(C (A_1 ^2 + A_2 ^2 + \dots + A_L ^2))$ $\leq O(C U) + O(C U' / C)^2$	$O(U) + O(R_{C-\{c_{i1}\}}) \leq O(U) + O(U' / C)^2$

5 Example

We use the decision system in table 2 as an example to illustrate the new algorithm, in order to illustrate the efficiency of the new algorithm better. For the fifteen objects in table 2, using Algorithm 1 to compute that $U' = \{X1, X2, X8, X4, X6, X7, X13\}$, $U/D = \{\{X1, X3, X4\}, \{X2, X5, X8, X12, X13\}, \{X6, X7, X10, X11, X15\}\} = \{D_1, D_2, D_3, D_4\}$

Table 2. The decision system

U	a	b	c	d	D
X1	1	1	1	1	0
X2	2	2	2	1	1
X3	1	1	1	1	0
X4	2	3	2	3	0
X5	2	2	2	1	1
X6	3	1	2	1	2
X7	1	2	3	2	2
X8	2	3	1	2	1
X9	3	1	2	1	3
X10	1	2	3	2	2
X11	3	1	2	1	2
X12	2	3	1	2	1
X13	4	3	4	2	1
X14	1	2	3	2	3
X15	4	3	4	2	2

According to the probability distribution function, we can obtain that $\mu_c(*)$ as follows: $\mu_c(X1) = (1, 0, 0, 0)$; $\mu_c(X2) = (0, 1, 0, 0)$; $\mu_c(X4) = (1, 0, 0, 0)$; $\mu_c(X6) = (0, 0, 2/3, 1/3)$; $\mu_c(X7) = (0, 0, 2/3, 1/3)$; $\mu_c(X8) = (0, 1, 0, 0)$; $\mu_c(X13) = (0, 1/2, 1/2, 0)$; The simplified decision system is shown as Table 3:

Table 3. Simplified decision system of Table 2

U	a	b	c	d	$\mu_c(*)$
X1	1	1	1	1	(1,0,0,0)
X2	2	2	2	1	(0,1,0,0)
X4	2	3	2	3	(1,0,0,0)
X6	3	1	2	1	(0,0,2/3,1/3)
X7	1	2	3	2	(0,0,2/3,1/3)
X8	2	3	1	2	(0,1,0,0)
X13	4	3	4	2	(0,1/2,1/2,0)

According to step 2, there is $M_1=4, M_2=3, M_3=4, M_4=2$, so $M_1 \geq M_3 \geq M_2 \geq M_4$ holds, the corresponding feature order is a, c, b, d . Select feature a as the first feature to feature selection, so there is $R = \{ a \}$; $R_c = \emptyset$. According to step 3, there is $U' / \{ a \} = \{ \{ X1, X7 \}, \{ X2, X4, X8 \}, \{ X6 \}, \{ X13 \} \}$. According to step 4, there

is $R_c = \{ \langle X1, X7 \rangle, \langle X2, X4 \rangle, \langle X8, X4 \rangle \}$. According to step 5, due to $R_c \neq \emptyset$, select $\{c\}$, then there is $R = \{a, c\}$, thus $R_c = \{ \langle X2, X4 \rangle \}$. Due to $R_c \neq \emptyset$, so select b , then $R = \{a, b, c\}$, thus $R_c = \emptyset$, the algorithm terminates, therefore, the feature selection $R = \{a, b, c\}$. The new algorithm is only to calculate four records, so as to reduce the computing time and the storage space, the performance of the new algorithm is better than other algorithms for feature selection.

5 Conclusion

To cut down the time complexity of feature selection algorithm based on conditional information entropy efficiently, on the basis of the simplified decision system, the definition of discernibility object pair set and the feature selection definition based on the discernibility object pair set are described, it is proved that the definition of feature selection is equal to the definition of feature selection based on conditional information entropy. On this condition, making use of the relations between the discernibility object pair and the feature selection, combining the idea of discernibility matrix, an efficient feature selection based on conditional information entropy is designed. In this paper, the better feature selection algorithm based on conditional information entropy is designed. This designing idea of this algorithm will provide a new thinking for other feature selection algorithms, which is our future work.

Acknowledgments. This work is supported by Education Department of Zhejiang Province (Grant No. Y201019091).

References

1. Pawlak, Z., Skowron, A.: Rudiments of rough sets. *Information Science* 117(1), 3–37 (2007)
2. Pawlak, Z., Wong, S.K., Ziarko, W.: Rough Sets: Probabilistic versus deterministic approach. *Computational Intelligence* 29(4), 81–95 (1988)
3. Zhang, W.X., Qiu, G.F., Wu, W.Z.: General theory of attribute reduction in rough set. *Science in China Series E* 35(12), 1304–1313 (2005)
4. Li, Y.S., Jia, R.Y.: Coverage Reduction algorithm based on conditional information entropy. *Computer Engineering* 36(16), 176–179 (2010)
5. Liu, Q.H., Li, F., Min, F.: An efficient knowledge reduction algorithm based on new conditional information entropy. *Control and Decision* 20(8), 878–882 (2005)
6. Jiang, Y., Wang, P.: Complete algorithm for attribute reduction based on discernibility matrix. *Computer Engineering and Applications* 43(19), 185–187 (2007)
7. Wang, J.Y., Gao, C.: Improved algorithm for attribute reduction based on discernibility matrix. *Computer Engineering* 35(3), 66–67 (2009)
8. Zhi, T.Y., Miao, D.Q.: The binary discernibility matrix transformation and high efficiency attributes reduction algorithms conformation. *Computer Science* 29(2), 140–142 (2003)
9. Xu, Z.Y., Yang, B.R., Song, W.: Quick algorithm for computing core based on discernibility object pair set. *Systems Engineering and Electronics* 30(4), 731–734 (2008)
10. Xu, Z.Y., Yang, B.R., Guo, Y.P.: Quick Algorithm for Computing Core Based on Information Entropy. *Journal of Chinese Computer Systems* 28(2), 279–282 (2007)

Author Index

- Baochun, Wu 446
Binbin, Gao 351
Bo, Jia 351
- Cao, Jin-Yan 377
Cao, Lai-Cheng 76
Cao, Yutao 84, 261
Chang, Weng-Long 483
Chen, Deng 438
Chen, Hui-Ying 135
Chen, Jiarui 100
Chen, Junjie 175, 183
Chen, Li 529
Chen, Shao 438
Chen, Yanmin 477
Cui, Jie 9
Cui, Xiaohong 175, 183
Cui, Yan-ping 199
- Deng, Xinguo 100
Ding, Huijie 430
Ding, Yu 544
Dong, Gang 147
Dong, Huajun 420
Duan, Yuexing 60
- Fan, Qing-Min 29
Feng, Yuqiang 318
Fu, Duan 392
Fu, Jia 155
Fujun, Feng 557
Fu-Qiang, Wang 141
- Gan, Yongquan 420
Gang, Li 280, 392
Gao, Li 147
Gao, Weizeng 1
Gao, Xin 147
Gong, Renxi 335
Gu, Hui 52
Gu, Meihua 310
Guo, Feipeng 230
Guo, Hongtao 161
Guo, Jun 370, 388
Guo, Xiaohong 357
- Guo, Yuming 161
Guodong, Li 302
- Han, Jiang-hong 34
Hao, Jinyan 310
Hao, Yonghua 113
He, Xiaojuan 161
Hu, Yi 245
Huang, Jincan 424
Huang, Min 570
Huo, Tieqiao 207
- Jia, Xiumei 295
Jia, Xiuming 399
Jianguo, Wang 590
Jin-xue, Zhang 325
Jin-zhong, Li 17
Jiyin, Zhao 446
Jun, Yu 120
Junshan, Li 557
- Li, Bing 454
Li, Hongbo 491
Li, Hua 286
Li, Hui 564
Li, Jiao 318
Li, Lin 499
Li, Qiang 399
Li, Wenhui 414
Li, Ye-li 521
Li, Yonggang 92
Li, Zhiqin 499
Li, Zhongxian 161
Liang, Yong-zhi 454
Liangjun, Wen 302
Lin, Daomiao 52
Lin, Xuan 364
Lin, Zhang 392
Liu, Chaofeng 382
Liu, Chun-tao 454
Liu, Jianguo 68
Liu, Li 414
Liu, Xiao-Feng 364
Liu, Yi 24
Liu, Yuxia 335

- Liu, Zhibin 550
 Liu, Zhong 424
 Lu, Aiping 161
 Luo, Jin'guang 214
 Luo, Jinkun 214
 Lv, Yong-Wei 364
- Ma, Xinming 24
 Ma, Yan 43
 Ma, Yanchun 537
 Ma, Yixuan 113
 Meng, Yu 414
 Miao, Guoyi 1
 Miao, Zhuang 529
 Min, Li 446
 Minzhi, Jia 351
- Na, Xue 113
 Niu, Xi-xian 199
- Pan, Dazhi 550
 Peng, Xinguang 253
 Peng, Zhiping 537
 Pengfei, Hu 302
 Ping, Xia 92
- Quanzhong, Ma 302
- Rong, Xing-fu 454
 Ruan, Jing 598
 Ruan, Ningjun 407
 Rui, Wang 280
 Ruirui, Zheng 446
- Sang, Lijun 183
 Shao-cong, Yan 267
 Shi, Guiming 420
 Shi, Lei 24
 Shuangyan, Hu 557
 Shu-ting, Zhang 267
 Song, Rugang 245
 Song, Zi-lin 529
 Su, Shoubao 43
 Sun, Lin 222
 Sun, Xihuan 84
- Tang, Hao 34
 Tian, Mu-Ling 370, 388
 Tian, Mu-Qin 135, 370, 377, 388
 Tie-Zhu, Qiao 141
- Wan, Shuting 92
 Wang, Ailian 60
 Wang, Ping 461
 Wang, Qianqian 343
 Wang, Qing-jiang 470
 Wang, Ranfeng 506
 Wang, Shu-Hua 364
 Wang, Xiao-dong 9
 Wang, Xiaoguo 222
 Wang, Ying 253
 Wang, Zhi-Peng 135
 Wei, Jinhong 343
 Wei, Sen-Sheng 370
 Wei-dong, Tang 17
 Wen, ZhiQiang 461
 Wu, Hongmei 414
 Wu, Jinping 245
 Wu, Suping 491
 Wu, Weiping 68
 Wu, Xiaojun 207
 Wu, Yunpeng 424
 Wu, Zhihui 461
- Xiangyang, Xu 582
 Xiao, Jihai 175, 183
 Xiao, Yu 128
 Xiao-Yu, Lu 141
 Xie, Kai 407
 Xingang, Wang 576
 Xiumei, Wei 576
 Xu, Huali 43
 Xue, Hua 128
 Xue-hua, Hu 17
 Xuesong, Jiang 576
 Xueying, Zhang 120
- Yan, Yaoxing 261
 Yang, Congyan 357
 Yang, Haifeng 239
 Yang, Yang 43, 582
 Yan-mei, Liu 267
 Yanxiao, Fu 582
 Yao, Lili 295
 Yao, Yangguang 100
 Yingjie, Zhang 302
 Yu, HuoQuan 407
 Yu, ShengHui 407
 Yu, Shuhao 43
 Yu, Zhou 1
 Yuan, Chang'an 214

- Yuan, Heqiang 343
Yuansheng, Li 582
Yufei, Wang 128
Yushui, Geng 576
- Zhan, Jingwen 477
Zhang, Changsheng 598
Zhang, Guojun 382
Zhang, Hong 230
Zhang, Junxi 207
Zhang, Li 564
Zhang, Lin 470
Zhang, Shouming 190
Zhang, Shu-Juan 514
Zhang, Weiming 424
- Zhang, Wen 407
Zhang, Yi 108
Zhang, Ying 529
Zhang, Yingmei 295
Zhao, Chen 92
Zhao, Jing 521
Zhao, Xiaofang 272
Zhao, Xiaoqiang 190
Zhao, Yao 382
Zheng, Guang 161
Zhou, Sen-xin 34
Zhu, Cheng 424
Zhu, Yanhui 461
Zou, Jiyan 420