

Fuchun Sun

Tianrui Li

Hongbo Li *Editors*

Foundations and Applications of Intelligent Systems

Proceedings of the Seventh International
Conference on Intelligent Systems and
Knowledge Engineering, Beijing, China,
Dec 2012 (ISKE 2012)



Springer

Advances in Intelligent Systems and Computing

Volume 213

Series Editor

J. Kacprzyk, Warsaw, Poland

For further volumes:

<http://www.springer.com/series/11156>

Fuchun Sun · Tianrui Li · Hongbo Li
Editors

Foundations and Applications of Intelligent Systems

Proceedings of the Seventh International
Conference on Intelligent Systems and
Knowledge Engineering, Beijing, China,
Dec 2012 (ISKE 2012)

 Springer

Editors

Fuchun Sun
Hongbo Li
Department of Computer Science
and Technology
Tsinghua University
Beijing
People's Republic of China

Tianrui Li
School of Information Science
and Technology
Southwest Jiaotong University
Chengdu
People's Republic of China

ISSN 2194-5357

ISSN 2194-5365 (electronic)

ISBN 978-3-642-37828-7

ISBN 978-3-642-37829-4 (eBook)

DOI 10.1007/978-3-642-37829-4

Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013947353

© Springer-Verlag Berlin Heidelberg 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law. The use of general descriptive names, registered names, trademarks, service marks, 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.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

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

Preface

This book is part of the Proceedings of the Seventh International Conference on Intelligent Systems and Knowledge Engineering (ISKE 2012) and the first International Conference on Cognitive Systems and Information Processing (CSIP 2012) held in Beijing, China, during December 15–17, 2012. ISKE is a prestigious annual conference on Intelligent Systems and Knowledge Engineering with past events held in Shanghai (2006, 2011), Chengdu (2007), Xiamen (2008), Hasselt, Belgium (2009), and Hangzhou (2010). Over the past few years, ISKE has matured into a well-established series of International Conferences on Intelligent Systems and Knowledge Engineering and related fields over the world. CSIP 2012 is the first conference sponsored by Tsinghua University and Science China Press, and technically sponsored by IEEE Computational Intelligence Society, Chinese Association for Artificial Intelligence. The aim of this conference is to bring together experts from different expertise areas to discuss the state of the art in cognitive systems and advanced information processing, and to present new research results and perspectives on future development. Both ISKE 2012 and CSIP 2012 provide academic forums for the participants to disseminate their new research findings and discuss emerging areas of research. It also creates a stimulating environment for the participants to interact and exchange information on future challenges and opportunities of Intelligent and Cognitive Science Research and Applications.

ISKE 2012 and CSIP 2012 received 406 submissions in total from about 1020 authors in 20 countries (United States of American, Singapore, Russian Federation, Saudi Arabia, Spain, Sudan, Sweden, Tunisia, United Kingdom, Portugal, Norway, Korea, Japan, Germany, Finland, France, China, Argentina, Australia, and Belgium). Based on rigorous reviews by the Program Committee members and reviewers, among 220 papers contributed to ISKE 2012, high-quality papers were selected for publication in the proceedings with the acceptance rate of 58.4 %. The papers were organized in 25 cohesive sections covering all major topics of Intelligent and Cognitive Science and Applications. In addition to the contributed papers, the technical program includes four plenary speeches by Jennie Si (Arizona State University, USA), Wei Li (California State University, USA), Chin-Teng Lin (National Chiao Tung University, Taiwan, China), and Guoqing Chen (Tsinghua University, China).

As organizers of both conferences, we are grateful to Tsinghua University, Science in China Press, Chinese Academy of Sciences for their sponsorship, grateful to IEEE Computational Intelligence Society, Chinese Association for Artificial Intelligence, State Key Laboratory on Complex Electronic System Simulation, Science and Technology on Integrated Information System Laboratory, Southwest Jiaotong University, University of Technology, Sydney, for their technical co-sponsorship.

We would also like to thank the members of the Advisory Committee for their guidance, the members of the International Program Committee and additional reviewers for reviewing the papers, and members of the Publications Committee for checking the accepted papers in a short period of time. Particularly, we are grateful to thank the publisher, Springer, for publishing the proceedings in the prestigious series of Advances in Intelligent Systems and Computing. Meanwhile, we wish to express our heartfelt appreciation to the plenary speakers, special session organizers, session chairs, and student volunteers. In addition, there are still many colleagues, associates, and friends who helped us in immeasurable ways. We are also grateful to them all. Last but not the least, we are thankful to all authors and participants for their great contributions that made ISKE 2012 and CSIP 2012 successful.

December 2012

Fuchun Sun
Tianrui Li
Hongbo Li

Organizing Committee

General Chair	Jie Lu (Australia) Fuchun Sun (China)
General Co-Chairs	Guoqing Chen (China) Yang Xu (China)
Honorary Chairs	L. A. Zadeh (USA) Bo Zhang (China)
Steering Committee Chairs	Etienne Kerre (Belgium) Zengqi Sun (China)
Organizing Chairs	Xiaohui Hu (China) Huaping Liu (China)
Program Chairs	Tianrui Li (China) Javier Montero (Spain)
Program Co-Chairs	Changwen Zheng (China) Luis Martínez López (Spain)
Sessions Chairs	Fei Song (China) Victoria Lopez (Spain)
Publications Chairs	Yuanqing Xia (China) Hongming Cai (China)
Publicity Chairs	Jiacun Wang (USA) Zheyang Zhang (Finland) Michael Sheng (Australia) Dacheng Tao (Australia)
Poster Chairs	Guangquan Zhang (Australia) Hongbo Li (China)

Members

Abdullah Al-Zoubi (Jordan)
 Andrzej Skowron (Poland)
 Athena Tocatlidou (Greece)
 B. Bouchon-Meunier (France)
 Benedetto Matarazzo (Italy)
 Bo Yuan (USA)
 Bo Zhang (China)
 Cengiz Kahraman (Turkey)
 Changwen Zheng (China)
 Chien-Chung Chan (USA)
 Cornelis Chris (Belgium)
 Dacheng Tao (Australia)
 Davide Ciucci (Italy)
 Davide Roverso (Norway)
 Du Zhang (USA)
 Enrico Zio (Italy)
 Enrique Herrera-Viedma (Spain)
 Erik Laes (Belgium)
 Etienne E. Kerre (Belgium)
 Francisco Chiclana (UK)
 Francisco Herrera (Spain)
 Fuchun Sun (China)
 Gabriella Pasi (Italy)
 Georg Peters (Germany)
 Germano Resconi (Italy)
 Guangquan Zhang (Australia)
 Guangtao Xue (China)
 Gulcin Buyukozkan (Turkey)
 Guolong Chen (China)
 Guoyin Wang (China)
 H.-J. Zimmermann (Germany)
 Huaping Liu (China)
 Hongbo Li (China)
 Hongjun Wang (China)
 Hongming Cai (China)
 Hongtao Lu (China)
 I. Burhan Turksen (Canada)
 Irina Perfilieva (Czech Republic)
 Jan Komorowski (Sweden)
 Janusz Kacprzyk (Poland)
 Javier Montero (Spain)
 Jer-Guang Hsieh (Taiwan China)

Michael Sheng (Australia)
 Mihir K. Chakraborty (India)
 Mike Nachtgeael (Belgium)
 Mikhail Moshkov (Russia)
 Min Liu (China)
 Peijun Guo (Japan)
 Pierre Kunsch (Belgium)
 Qi Wang (China)
 Qingsheng Ren (China)
 Rafael Bello (Cuba)
 Richard Jensen (UK)
 Ronald R. Yager (USA)
 Ronei Marcos de Moraes (Brasil)
 Ryszard Janicki (Canada)
 S. K. Michael Wong (Canada)
 Shaojie Qiao (China)
 Shaozi Li (China)
 Sheela Ramanna (Canada)
 Su-Cheng Haw (Malaysia)
 Suman Rao (India)
 Sushmita Mitra (India)
 Takehisa Onisawa (Japan)
 Tetsuya Murai (Japan)
 Tianrui Li (China)
 Tzung-Pei Hong (Taiwan, China)
 Ufuk Cebeci (Turkey)
 Victoria Lopez (Spain)
 Vilem Novak (Czech Republic)
 Weiming Shen (Canada)
 Weixing Zhu (China)
 Wensheng Zhang (China)
 Witold Pedrycz (Canada)
 Wujun Li (China)
 Xiaohui Hu (China)
 Xianyi Zeng (France)
 Xiaogang Jin (China)
 Xiaoqiang Lu (China)
 Xiaoyan Zhu (China)
 Xiao-Zhi Gao (Finland)
 Xuelong Li (China)
 Xun Gong (China)
 Yan Yang (China)

Jesús Vega (Spain)	Yangguang Liu (China)
Jiacun Wang (USA)	Yanmin Zhu (China)
Jianbo Yang (UK)	Yaochu Jin (Germany)
Jie Lu (Australia)	Yasuo Kudo (Japan)
Jingcheng Wang (China)	Yi Tang (China)
Jitender S. Deogun (USA)	Yinglin Wang (China)
Jouni Jarvinen (Finland)	Yiyu Yao (Canada)
Juan-Carlos Cubero (Spain)	Yongjun Shen (Belgium)
Jun Liu (UK)	Yuancheng Huang (China)
Jyrki Nummenmaa (Finland)	Yuanqing Xia (China)
Koen Vanhoof (Belgium)	Zbigniew Suraj (Poland)
Krassimir Markov (Bulgaria)	Zbigniew W. Ras (U.S.A)
Liliane Santos Machado (Brasil)	Zengqi Sun (China)
Lisheng Hu (China)	Zheyang Zhang (Finland)
Luis Magdalena (Spain)	Zhong Li (Germany)
Luis Martinez López (Spain)	Zhongjun He (China)
Lusine Mkrtychyan (Italy)	Zhongzhi Shi (China)
Madan M. Gupta (Canada)	Bo Peng (China)
Martine De Cock (Belgium)	Fei Teng (China)
Masoud Nikravesht (USA)	

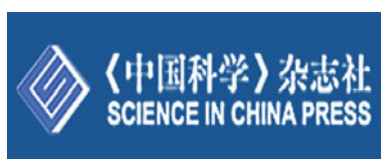
Sponsors



Tsinghua University



Chinese Academy of Sciences



Science in China Press



Institute of Electrical and Electronics Engineers

Contents

The Incremental Updating Method for Core Computing Based on Information Entropy in Set-Valued Ordered Information Systems	1
Chuan Luo, Tianrui Li, Hongmei Chen and Shaoyong Li	
An Algebraic Analysis for Binary Intuitionistic L-Fuzzy Relations . . .	11
Xiaodong Pan and Peng Xu	
An Improved Classification Method Based on Random Forest for Medical Image.	21
Niu Zhang, Haiwei Pan, Qilong Han and Xiaoning Feng	
IdeaGraph: Turning Data into Human Insights for Collective Intelligence	33
Hao Wang, Yukio Ohsawa, Pin Lv, Xiaohui Hu and Fanjiang Xu	
A General Hierarchy-Set-Based Indoor Location Modeling Approach.	45
Jingyu Sun, Huifang Li and Hong Huang	
Acquisition of Class Attributes Based on Chinese Online Encyclopedia and Association Rule Mining	61
Zhen Jia, Hongfeng Yin and Dake He	
A Delivery Time Computational Model for Pervasive Computing in Satellite Networks	73
Jianzhou Chen, Lixiang Liu and Xiaohui Hu	
How People Use Visual Landmarks for Locomotion Distance Estimation: The Case Study of Eye Tracking	87
Huiting Zhang and Kan Zhang	

Prediction Analysis of the Railway Track State Based on PCA-RBF Neural Network	99
Yan Yang, Xuyao Lu, Qi Dai and Hongjun Wang	
A Two-Step Agglomerative Hierarchical Clustering Method for Patent Time-Dependent Data	111
Hongshu Chen, Guangquan Zhang, Jie Lu and Donghua Zhu	
A Novel Modular Recurrent Wavelet Neural Network and Its Application to Nonlinear System Identification	123
Haiquan Zhao and Xiangping Zeng	
Automated Text Data Extraction Based on Unsupervised Small Sample Learning	133
Yulong Liu, Shengsheng Shi, Chunfeng Yuan and Yihua Huang	
SLAM and Navigation of a Mobile Robot for Indoor Environments	151
Shuangshuang Lei and Zhijun Li	
A Novel Approach for Computing Quality Map of Visual Information Fidelity Index	163
Yu Shao, Fuchun Sun, Hongbo Li and Ying Liu	
A Simulation of Electromagnetic Compatibility QoS Model in Cognitive Radio Network	175
Hai-yu Ren, Ming-xue Liao, Xiao-xin He and Kun Xu	
Separating and Recognizing Gestural Strokes for Sketch-Based Interfaces	187
Yougen Zhang, Hanchen Song, Wei Deng and Lingda Wu	
Modeling and Performance Analysis of Workflow Based on Advanced Fuzzy Timing Petri Nets	199
Huifang Li and Xinfang Cui	
Price Discount Subsidy Contract for a Two-Echelon Supply Chain with Random Yield	213
Weimin Ma, Xiaoxi Zhu and Miaomiao Wang	
Reducing EMI in a PC Power Supply with Chaos Control	231
Yuhong Song, Zhong Li, Junying Niu, Guidong Zhang, Wolfgang Halang and Holger Hirsch	

Optimization of Plasma Fabric Surface Treatment by Modeling with Neural Networks 243
 Radhia Abd Jelil, Xianyi Zeng, Ludovic Koehl and Anne Perwuelz

Model-Driven Approach for the Development of Web Application System 257
 Jinkui Hou

A Web UI Modeling Approach Supporting Model-Driven Software Development. 265
 Jinkui Hou

On the Possibilistic Handling of Priorities in Access Control Models 275
 Salem Benferhat, Khalid Bouriche and Mohamed Ouzarf

Fractional Order Control for Hydraulic Turbine System Based on Nonlinear Model. 287
 Xinjian Yuan

Fingerprint Orientation Estimation Based on Tri-Line Model. 297
 Xiaolong Zheng, Canping Zhu and Jicheng Meng

Joint Rate and Power Allocation for Cognitive Radios Networks with Uncertain Channel Gains 309
 Zhixin Liu, Jinfeng Wang, Hongjiu Yang and Kai Ma

Link Prediction Based on Sequential Bayesian Updating in a Terrorist Network 321
 Cheng Jiang, Juyun Wang and Hua Yu

Comparison Between Predictive and Predictive Fuzzy Controller for the DC Motor via Network 335
 Abdallah A. Ahmed and Yuanqing Xia

Trajectory Tracking Method for UAV Based on Intelligent Adaptive Control and Dynamic Inversion. 347
 Juan Dai and Yuanqing Xia

Evolutionary Game Analysis on Enterprise’s Knowledge-Sharing in the Cooperative Networks 359
 Shengliang Zong, Zhen Cai and Mingyue Qi

An Ant Colony Algorithm for Two-Sided Assembly Line Balancing Problem Type-II 369
 Zeqiang Zhang, Junyi Hu and Wenming Cheng

Detection of Earnings Manipulation with Multiple Fuzzy Rules 379
 Shuangjie Li and Hongxu Liang

A New Unbalanced Linguistic Scale for the Classification of Olive Oil Based on the Fuzzy Linguistic Approach. 389
 M. Espinilla, F. J. Estrella and L. Martínez

Algorithm to Find Ground Instances in Linguistic Truth-Valued Lattice-Valued First-Order Logic $\mathcal{L}_{V(n \times 2)}F(X)$ 401
 Xiaomei Zhong, Yang Xu and Peng Xu

Estimating Online Review Helpfulness with Probabilistic Distribution and Confidence 411
 Zunqiang Zhang, Qiang Wei and Guoqing Chen

Probabilistic Attribute Mapping for Cold-Start Recommendation 421
 Guangxin Wang and Yinglin Wang

Measuring Landscapes Quality Using Fuzzy Logic and GIS 433
 Victor Estévez González, Luis Garmendia Salvador and Victoria López López

Ship Dynamic Positioning Decoupling Control Based on ADRC 443
 Zhengling Lei, Guo Chen and Liu Yang

Intelligent Double-Eccentric Disc Normal Adjustment Cell in Robotic Drilling 457
 Peijiang Yuan, Maozhen Gong, Tianmiao Wang, Fucun Ma, Qishen Wang, Jian Guo and Dongdong Chen

A Method of Fuzzy Attribute Reduction Based on Covering Information System 469
 Fachao Li, Jiaoying Wang and Chenxia Jin

Sliding Mode-Like Fuzzy Logic Control with Boundary Layer Self-Tuning for Discrete Nonlinear Systems 479
 Xiaoyu Zhang and Fang Guo

Design of Direct Adaptive Fuzzy Sliding Mode Control for Discrete Nonlinear System 491
 Xiping Zhang and Xiaoyu Zhang

Application of Improved Particle Swarm Optimization-Neural Network in Long-Term Load Forecasting. 503
 Xuelian Gao, Yanyu Chen, Zhennan Cui, Nan Feng, Xiaoyu Zhang and Lei Zhao

Routing Techniques Based on Swarm Intelligence 515
 Delfín Rupérez Cañas, Ana Lucila Sandoval Orozco and Luis Javier García Villalba

Bayesian Regularization BP Neural Network Model for the Stock Price Prediction 521
 Qi Sun, Wen-Gang Che and Hong-Liang Wang

Adaptive Fuzzy Control for Wheeled Mobile Manipulators with Switching Joints. 533
 Zhijun Li

Bundle Branch Blocks Classification Via ECG Using MLP Neural Networks 547
 Javier F. Fornari and José I. Peláez Sánchez

Fuzzy Associative Classifier for Probabilistic Numerical Data. 563
 Bin Pei, Tingting Zhao, Suyun Zhao and Hong Chen

Adaptive Boosting for Enhanced Vortex Visualization 579
 Li Zhang, Raghu Machiraju, David Thompson, Anand Rangarajan and Xiangxu Meng

Applying Subgroup Discovery Based on Evolutionary Fuzzy Systems for Web Usage Mining in E-Commerce: A Case Study on OrOliveSur.com 591
 C. J. Carmona, M. J. del Jesus and S. García

Distributed Data Distribution Mechanism in Social Network Based on Fuzzy Clustering 603
 Yan Cao, Jian Cao and Minglu Li

Experimental Comparisons of Instances Set Reduction Algorithms 621
 Yuelin Yu, Yangguang Liu, Bin Xu and Xiaoqi He

Recycla.me: Technologies for Domestic Recycling 631
 Guadalupe Miñana, Victoria Lopez and Juan Tejada

Caching Mechanism in Publish/Subscribe Network 641
 Hongjie Tan, Jian Cao and Minglu Li

**Vision-Based Environmental Perception and Navigation
 of Micro-Intelligent Vehicles** 653
 Ming Yang, Zhengchen Lu, Lindong Guo, Bing Wang
 and Chunxiang Wang

**Dynamic Surface Control of Hypersonic Aircraft
 with Parameter Estimation** 667
 Bin Xu, Fuchun Sun, Shixing Wang and Hao Wu

**Nonlinear Optimal Control for Robot Manipulator
 Trajectory Tracking** 679
 Shijie Zhang, Ning Yi and Fengzhi Huang

**Impedance Identification and Adaptive Control of Rehabilitation
 Robot for Upper-Limb Passive Training** 691
 Aiguo Song, Lizheng Pan, Guozheng Xu and Huijun Li

Independent Component Analysis: Embedded LTSA 711
 Qianwen Yang, Yuan Li, Fuchun Sun and Qingwen Yang

**Analysis and Improvement of Anonymous Authentication
 Protocol for Low-Cost RFID Systems** 723
 Zhijun Ge and Yongsheng Hao

**Formation Control and Stability Analysis of Spacecraft:
 An Energy Concept-Based Approach** 733
 Zhijie Gao, Fuchun Sun, Tieding Guo, Haibo Min
 and Dongfang Yang

An Intelligent Conflict Resolution Algorithm of Multiple Airplanes . . . 745
 Jingjuan Zhang, Jun Wu and Rong Zhao

**A Novel End-to-End Authentication Protocol for Satellite
 Mobile Communication Networks** 755
 Xiaoliang Zhang, Heyu Liu, Yong Lu and Fuchun Sun

Current Status, Challenges and Outlook of E-health Record Systems in China 767
Xiangzhu Gao, Jun Xu, Golam Sorwar and Peter Croll

Singularity Analysis of the Redundant Robot with the Structure of Three Consecutive Parallel Axes 779
Gang Chen, Long Zhang, Qingxuan Jia and Hanxu Sun

Statistical Analysis of Stock Index Return Rate Distribution in Shanghai and Shenzhen Stock Market 793
Guan Li

The Incremental Updating Method for Core Computing Based on Information Entropy in Set-Valued Ordered Information Systems

Chuan Luo, Tianrui Li, Hongmei Chen and Shaoyong Li

Abstract Attribute core is an important concept of attribute reduction in rough set theory. In this paper, we focus on analyzing the incremental updating method for core computing based on information entropy in set-valued ordered information systems. The definition of attribute core by means of information entropy in set-valued ordered information system is introduced. And the changing mechanisms of information entropy and attribute core are analyzed while the object set varies with time. On the basis of the changing mechanisms, the attribute core can be updated incrementally based on existing results to reduce redundant computation. Finally, an example illustrates the validity of the proposed approach.

Keywords Rough set theory · Attribute core · Incremental · Information entropy · Set-valued ordered information systems

1 Introduction

The rough set theory (RST), proposed by Pawlak [1], is toward to complete information system initially. However, in many practical issues, it may happen that some of the attribute values for an object are missing, and these missing values can

C. Luo · T. Li (✉) · H. Chen · S. Li
School of Information Science and Technology, Southwest Jiaotong University,
610031 Chengdu, China
e-mail: trli@swjtu.edu.cn

C. Luo
e-mail: luochuan@my.swjtu.edu.cn

H. Chen
e-mail: hmchen@swjtu.edu.cn

S. Li
e-mail: meterer@163.com

be represented by the set of all possible values for the attribute or equivalence by the domain of the attribute. Set-valued information system is a generalized model of single-valued information system, which can be used to characterize uncertain information and missing information in information systems [2, 3]. Incomplete information system can also be regarded as a special case of set-valued information system (disjunctively interpreted set-valued information system). On the other hand, attributes in the information systems are sometimes with preference-ordered domains, which the ordering of properties of attributes plays a crucial role. For this reason, Greco et al. proposed an extension RST, called Dominance-based Rough Sets Approach (DRSA) to take into account the ordering properties of attributes [4, 5]. Presently, several further studies have been made about properties and algorithmic implementations of DRSA [3, 6, 7, 18].

With the growing volume of data, it often occurs that the information systems processed by RST have millions of data records, and the number of records increases dynamically. Most of traditional knowledge acquisition algorithms based on RST are designed for static data processing, but many real data in the information systems are collected dynamically. So, these static algorithms will be ineffective in such situations. Thus, designing an incremental algorithm for knowledge acquisition is desirable [8, 9]. Now, there has been much research on incremental updating in RST, such as incremental rule extraction [10], incremental updating approximations [16], incremental attribute reduction [11], and incremental updating attribute core [12]. In this paper, we focus on the problem of incrementally computing attribute core from the dynamic information systems.

Attribute reduction is a focus issue in RST and its applications. In the last twenty years, many attribute reduction methods based on RST have been developed. However, most of algorithms for attribute reduction proposed are on the basis of the determination of attribute core. The computation of attribute core has become a key step in the solution of the problem of attribute reduction. Then, designing an effective approach for core computing has important practical value [13]. Since the theory of entropy has first been introduced to measure the uncertainty of the structure of a system by Shannon in 1948, many researchers have attempted to use the entropy to measure the uncertainty in RST [14]. They also considered adopting the entropy as the heuristic metrics to achieve core and reduction in attribute effectively. Wang et al. [13] proposed an algorithm for calculating the attribute core of a decision table based on information entropy. Xu et al. [15] defined a simplified decision table and proved that the core calculated based on information entropy in the simplified decision table is the same to the original decision table. Then, an improved algorithm for computing attribute core was proposed. On the other hand, many real data in the information systems are collected dynamically. Thus, it is desirable to design an incremental algorithm for knowledge acquisition [16–19]. Yang et al. [20] introduced an incremental updating algorithm for computation of a core based on the discernibility matrix in dynamic computation. Jia et al. [21] redefined the dominance discernibility matrix to analyze incremental updating for core computation in the DRSA. Liang [22] also analyzed the changing mechanism of information entropy when a new object

is inserted into the decision system and developed an incremental algorithm for core computing. In this paper, we focus on the incremental updating method for computation of a core based on information entropy in set-valued ordered information systems (SOIS).

The paper is organized as follows. In Sect. 1, some basic notions about SOIS are introduced. The definition of attribute core in the SOIS is also presented from the information view. In Sect. 2, the changing mechanism of information entropy and the principle of updating core in SOIS are discussed while the object set varies. An incremental algorithm for updating core is proposed in SOIS when inserting an object into the universe in Sect. 3. In Sect. 4, we illustrate how to update core in SOIS based on information entropy when the object set evolves over time. Finally, we conclude this paper in Sect. 5.

2 Preliminaries

In this section, several basic concepts and preliminaries of SOIS are introduced here firstly [2, 3].

2.1 Set-Valued Ordered Information Systems

A set-valued information system is an ordered quadruple (U, AT, V, f) , where $U = \{x_1, x_2, \dots, x_n\}$ is a non-empty finite set of objects, $AT = \{a_1, a_2, \dots, a_l\}$ is a finite set of attributes, $V = \{V_{a_1}, V_{a_2}, \dots, V_{a_l}, a_i \in AT\}$ is the set of attribute values, $V_{a_i} (i = 1, 2, \dots, l)$ is domain of attribute $a_i (a_i \in AT)$. f is a mapping from $U \times AT$ to V such that $f : U \times AT \rightarrow 2^V$ is a set-valued mapping.

In an information system, if the domain of an attribute values is ordered according to a decreasing or increasing preference, then the attribute is a criterion.

Definition 1 A set-valued information system is called a SOIS if all attributes are criterions.

Definition 2 Let (U, AT, V, f) be a SOIS and $A \subseteq AT$. The dominance relation in terms of A is defined as:

$$R_A^{\supseteq} = \{(x, y) \in U \times U \mid f(y, a) \supseteq f(x, a) (\forall a \in A)\} \quad (1)$$

The dominance class induced by the dominance relation R_A^{\supseteq} are the set of objects may dominating x , i.e.,

$$[x]_A^{\supseteq} = \{y \in U \mid f(y, a) \supseteq f(x, a) (\forall a \in A)\} \quad (2)$$

The set of objects may be dominated by x is denoted as follows.

$$[x]_A^{\subseteq} = \{y \in U \mid f(y, a) \subseteq f(x, a) (\forall a \in A)\} \quad (3)$$

where $[x]_A^{\supseteq}$ and $[x]_A^{\subseteq}$ are called the A -dominating set and A -dominated set with respect to $x \in U$ in SOIS, respectively.

Definition 3 Let $S = (U, AT, V, f)$ be a set-valued ordered information system, for any $X \subseteq U$ and $A \subseteq AT$, the lower and the upper approximation of X with respect to the dominance relation R_A^{\supseteq} are defined as follows:

$$\underline{R}_A^{\supseteq}(X) = \{x \in U \mid [x]_A^{\supseteq} \subseteq X\}; \quad (4)$$

$$\overline{R}_A^{\supseteq}(X) = \{x \in U \mid [x]_A^{\supseteq} \cap X \neq \emptyset\}. \quad (5)$$

2.2 The Attribute Core in SOIS from the Information View

In the following, we will introduce the concept of information entropy in the SOIS and present the definition of attribute core by means of information entropy.

Definition 4 Let (U, AT, V, f) be a SOIS and $A \subseteq AT$. The information entropy of knowledge A is defined as

$$I(A) = \frac{1}{|U|} \sum_{i=1}^{|U|} \left(1 - \frac{|[x_i]_A^{\supseteq}|}{|U|}\right) = 1 - \frac{1}{|U|^2} \sum_{i=1}^{|U|} |[x_i]_A^{\supseteq}| \quad (6)$$

where $|\cdot|$ denotes the cardinality of a set.

Proposition 1 Let (U, AT, V, f) be a SOIS and $A, B \subseteq AT$. If $B \subseteq A$, then $I(A) \geq I(B)$.

Proposition 2 Let (U, AT, V, f) be a SOIS. The attribute $a \in AT$ is indispensable if and only if $I(AT) - I(AT - \{a\}) > 0$.

Definition 5 Let (U, AT, V, f) be a SOIS and $A \subseteq AT$. If $\forall a \in A$ satisfies $I(AT) - I(AT - \{a\}) > 0$, then A is called the attribute core and denoted by $Core(AT)$.

3 Principles for Incrementally Updating Core in SOIS Based on Information Entropy

In this section, we analyze principles for incrementally updating information entropy and attribute core while the object set varies with time. For simplicity, we assume that $I(A)$ is information entropy of the knowledge A in the original SOIS,

and $Core(AT)$ is the original attribute core of the SOIS. Let $I'(A)$ denote information entropy of knowledge A and $Core'(AT)$ as the attribute core when the object \tilde{x} is inserted into or deleted from the information system.

Definition 6 Let (U, AT, V, f) be a SOIS and $A \subseteq AT$. The definite dominance relation in terms of A is defined as:

$$R_A^{\supseteq} = \{(x, y) \in U \times U \mid f(y, a) \supseteq f(x, a) (\forall a \in A)\} \quad (7)$$

The dominance class induced by the definite dominance relation R_A^{\supseteq} are the set of objects dominating x definitely, i.e.,

$$[x]_A^{\supseteq} = \{y \in U \mid f(y, a) \supseteq f(x, a) (\forall a \in A)\} \quad (8)$$

The set of objects dominated by x definitely is denoted as follows.

$$[x]_A^{\subseteq} = \{y \in U \mid f(y, a) \subseteq f(x, a) (\forall a \in A)\} \quad (9)$$

where $[x]_A^{\supseteq}$ and $[x]_A^{\subseteq}$ are called the definite A -dominating set and definite A -dominated set with respect to $x \in U$ in SOIS, respectively.

It is easy to see that $[x]_A^{\supseteq} = [x]_A^{\supseteq} - \{x\}$, $[x]_A^{\subseteq} = [x]_A^{\subseteq} - \{x\}$.

3.1 Insertion of a New Object

Proposition 3 Let (U, AT, V, f) be a SOIS and $A \subseteq AT$. After adding a new object \tilde{x} to the information system, we have:

$$I'(A) = 1 - \frac{1}{(|U| + 1)^2} (|U|^2(1 - I(A)) + |[[\tilde{x}]_A^{\supseteq}]| + |[[\tilde{x}]_A^{\subseteq}]|) \quad (10)$$

where $[[\tilde{x}]_A^{\supseteq}]$ and $[[\tilde{x}]_A^{\subseteq}]$ are the A -dominating set and definite A -dominated set with respect to \tilde{x} , respectively.

Proof Let $\rho_i = [x_i]_A^{\supseteq}$, $1 \leq i \leq |U|$ be the A -dominating set with respect to x_i in the original information system and $\rho'_i = [x_i]_A^{\supseteq}$, $1 \leq i \leq |U| + 1$ be the A -dominating set respect to x_i after the new object \tilde{x} is added to the information system.

$$\begin{aligned} I'(A) &= 1 - \frac{1}{(|U| + 1)^2} \left(\sum_{i=1}^{|U|+1} |[[\tilde{x}]_A^{\supseteq}]| \right) \\ &= 1 - \frac{1}{(|U| + 1)^2} \left(\sum_{i=1}^{|U|} |\rho'_i| + |\rho'_{|U|+1}| \right) \end{aligned}$$

$$\begin{aligned}
&= 1 - \frac{1}{(|U| + 1)^2} \left(\sum_{i=1}^{|U|} |\rho_i| + |\rho'_{|U|+1}| + \left(\sum_{i=1}^{|U|} |\rho'_i| - \sum_{i=1}^{|U|} |\rho_i| \right) \right) \\
&= 1 - \frac{1}{(|U| + 1)^2} \left(\sum_{i=1}^{|U|} |\rho_i| + |\rho'_{|U|+1}| + |[\tilde{x}]_A^c| \right) \\
&= 1 - \frac{1}{(|U| + 1)^2} \left(|U|^2 \left(1 - \left(1 - \frac{1}{|U|^2} \sum_{i=1}^{|U|} |\rho_i| \right) \right) + |\rho'_{|U|+1}| + |[\tilde{x}]_A^c| \right) \\
&= 1 - \frac{1}{(|U| + 1)^2} \left(|U|^2(1 - I(A)) + |[\tilde{x}]_A^{\supseteq}| + |[\tilde{x}]_A^c| \right)
\end{aligned}$$

Proposition 4 *Let (U, AT, V, f) be a SOIS and $B \subseteq A \subseteq AT$. If $I(A) > I(B)$, then we have that $I'(A) > I'(B)$ after inserting a new object \tilde{x} into U .*

Proof When the new object \tilde{x} is inserted into U , $U' = U \cup \{\tilde{x}\}$. From the definitions of A -dominating set $[x]_A^{\supseteq}$ and the definite A -dominated set $[x]_A^c$, it is easy to obtain that if $B \subseteq A$, then $\forall x \in U'$, we have $[x]_B^{\supseteq} \supseteq [x]_A^{\supseteq}$ and $[x]_B^c \supseteq [x]_A^c$, i.e., $|[x]_B^{\supseteq}| \geq |[x]_A^{\supseteq}|$ and $|[x]_B^c| \geq |[x]_A^c|$. Since the assumption of $I(A) > I(B)$, there exists that $\exists x \in U$, $|[x]_B^{\supseteq}| > |[x]_A^{\supseteq}|$. Therefore, based on the definition of information entropy, it is easy to see that $I'(A) > I'(B)$.

Proposition 5 *Let (U, AT, V, f) be a SOIS. If $\text{Core}(AT)$ is the core of the original information system, and $\text{Core}'(AT)$ is the core after the new object \tilde{x} is inserted into the information system, then we have that $\text{Core}'(AT) \supseteq \text{Core}(AT)$.*

Proof Since $\text{Core}(AT)$ is the attribute core of the original information system, then $\forall a \in \text{Core}(AT)$, $I(AT) > I(AT - \{a\})$. Therefore, after inserting the new object \tilde{x} into U , we have that $I'(AT) > I'(AT - \{a\})$ by Proposition 3, i.e., $\forall a \in \text{Core}(AT)$, $I'(AT) > I'(AT - \{a\})$. From the definition of the attribute core, we obtain that $\text{Core}'(AT) \supseteq \text{Core}(AT)$.

3.2 Deletion of an Object

Proposition 6 *Let (U, AT, V, f) be a SOIS and $A \subseteq AT$. After deleting an object \tilde{x} from the information system, we have:*

$$I'(A) = 1 - \frac{1}{(|U| - 1)^2} \left(|U|^2(1 - I(A)) - |[\tilde{x}]_A^{\supseteq}| - |[\tilde{x}]_A^c| \right) \quad (11)$$

where $[\tilde{x}]_A^{\supseteq}$ and $[\tilde{x}]_A^c$ are the A -dominating set and definite A -dominated set with respect to \tilde{x} , respectively.

Proof Let $\rho_i = [x_i]_A^{\supseteq}$, $\mu_j = [x_j]_A^{\subsetneq}$, $1 \leq i \leq |U|$ be the A -dominating set with respect to x_i in the original information system and $\rho'_i = [x_i]_A^{\supseteq}$, $1 \leq i \leq |U| - 1$ be the A -dominating set respect to x_i after the object \tilde{x} is deleted from the information system.

$$\begin{aligned}
I'(A) &= 1 - \frac{1}{(|U| - 1)^2} \left(\sum_{i=1}^{|U|-1} |\rho'_i| \right) \\
&= 1 - \frac{1}{(|U| - 1)^2} \left(\sum_{i=1}^{|U|} |\rho_i| - |\rho_j| - \left(\left(\sum_{i=1}^{|U|} |\rho_i| - |\rho_j| \right) - \sum_{i=1}^{|U|-1} |\rho'_i| \right) \right) \\
&= 1 - \frac{1}{(|U| - 1)^2} \left(|U|^2 \left(1 - \left(1 - \frac{1}{|U|^2} \sum_{i=1}^{|U|} |\rho_i| \right) \right) - |\rho_j| - |\mu_j| \right) \\
&= 1 - \frac{1}{(|U| - 1)^2} \left(|U|^2 (1 - I(A)) - |[x]_A^{\supseteq}| - |[x]_A^{\subsetneq}| \right)
\end{aligned}$$

Proposition 7 Let (U, AT, V, f) be a SOIS and $B \subseteq A \subseteq AT$. When an object \tilde{x} is deleted from U , we have that: if $I'(A) > I'(B)$, then $I(A) > I(B)$.

Proof When the object \tilde{x} is deleted from U , $U' = U - \{\tilde{x}\}$. From the definitions of A -dominating set $[x]_A^{\supseteq}$ and the definite A -dominated set $[x]_A^{\subsetneq}$, it is easy to obtain that if $B \subseteq A$, then $\forall x \in U'$, we have $[x]_B^{\supseteq} \supseteq [x]_A^{\supseteq}$ and $[x]_B^{\subsetneq} \supseteq [x]_A^{\subsetneq}$, i.e., $|[x]_B^{\supseteq}| \geq |[x]_A^{\supseteq}|$ and $|[x]_B^{\subsetneq}| \geq |[x]_A^{\subsetneq}|$. Since the assumption of $I'(A) > I'(B)$, there exists that $\exists x \in U'$, $|[x]_B^{\supseteq}| > |[x]_A^{\supseteq}|$. Therefore, based on the definition of information entropy, it is easy to obtain that $I(A) > I(B)$.

Proposition 8 Let (U, AT, V, f) be a SOIS. If $\text{Core}(AT)$ is the core of the original information system, and $\text{Core}'(AT)$ is the core after the object \tilde{x} is deleted from the information system, then we have $\text{Core}(AT) \supseteq \text{Core}'(AT)$.

Proof Since $\text{Core}'(AT)$ is the attribute core of the updated information system when the object \tilde{x} is deleted from the universe, then $\forall a \in \text{Core}'(AT)$, we have $I'(AT) > I'(AT - \{a\})$. From the Proposition 5, if $I'(AT) > I'(AT - \{a\})$, then $I(AT) > I(AT - \{a\})$. Hence, we can see that $\forall a \in \text{Core}'(AT)$, $I(AT) > I(AT - \{a\})$. Therefore, from the definition of the attribute core, we have $\text{Core}(AT) \supseteq \text{Core}'(AT)$.

4 A Illustrate Example

In this section, we use an example to illustrate that how to incrementally updating attribute core in SOIS based on information entropy when the object set varies over time. Considering the SOIS $S = (U, AT, V, f)$ as given in Tables 1 and 2,

where $U = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ in Table 1 and $U = \{x_7\}$ in Table 2, $AT = \{a_1, a_2, a_3, a_4\} = \{\text{Audition, Spoken language, Reading, Writing}\}$ and $V = \{E, F, G\} = \{\text{English, French, German}\}$.

Now, we consider the following two cases:

1. The object x_7 in Table 2 is inserted to the SOIS as showed in Table 1;
2. The object x_6 is deleted from the SOIS as showed in Table 1.

Firstly, we use the static method to achieve the attribute core of the original information system (see Table 1). We have the following results:

$$\begin{aligned} U/R_{AT}^{\supseteq} &= \{\{x_1, x_2, x_3\}, \{x_2\}, \{x_2, x_3\}, \{x_4\}, \{x_5\}, \{x_2, x_6\}\}; \\ U/R_{AT-\{a_1\}}^{\supseteq} &= \{\{x_1, x_2, x_3\}, \{x_2\}, \{x_2, x_3\}, \{x_4\}, \{x_5\}, \{x_2, x_6\}\}; \\ U/R_{AT-\{a_2\}}^{\supseteq} &= \{\{x_1, x_2, x_3\}, \{x_2\}, \{x_2, x_3\}, \{x_4\}, \{x_5\}, \{x_2, x_6\}\}; \\ U/R_{AT-\{a_3\}}^{\supseteq} &= \{\{x_1, x_2, x_3, x_4\}, \{x_2, x_4\}, \{x_2, x_3, x_4\}, \{x_2, x_4\}, \{x_2, x_4, x_5\}, \\ & x_2, x_4, x_6\}\}; \\ U/R_{AT-\{a_4\}}^{\supseteq} &= \{\{x_1, x_2, x_3, x_6\}, \{x_2\}, \{x_2, x_3\}, \{x_4\}, \{x_5\}, \{x_2, x_6\}\}. \end{aligned}$$

Then, we calculate the information entropy of the attribute set by Definition 4:
 $I(AT) = \frac{13}{18}$, $I(AT - \{a_1\}) = \frac{13}{18}$, $I(AT - \{a_2\}) = \frac{13}{18}$, $I(AT - \{a_3\}) = \frac{19}{36}$,
 $I(AT - \{a_4\}) = \frac{25}{36}$.

Finally, we have $Core(AT) = \{a_3, a_4\}$ in Table 1 from Definition 5.

Now, we consider the two cases as above, and compute the attribute core according to the incremental updating principles:

- When the object x_7 in Table 2 is inserted into Table 1, the attribute core is updated by as follows:

- (a) $[x_7]_{AT}^{\supseteq} = \{x_2, x_7\}$; $[x_7]_{AT}^{\subsetneq} = \emptyset$; $[x_7]_{AT-\{a_1\}}^{\supseteq} = \{x_2, x_3, x_7\}$; $[x_7]_{AT-\{a_1\}}^{\subsetneq} = \emptyset$;
 $[x_7]_{AT-\{a_2\}}^{\supseteq} = \{x_2, x_7\}$; $[x_7]_{AT-\{a_2\}}^{\subsetneq} = \emptyset$; $[x_7]_{AT-\{a_3\}}^{\supseteq} = \{x_2, x_4, x_7\}$;
 $[x_7]_{AT-\{a_3\}}^{\subsetneq} = \{x_1, x_3, x_7\}$; $[x_7]_{AT-\{a_4\}}^{\supseteq} = \{x_2, x_7\}$; $[x_7]_{AT-\{a_4\}}^{\subsetneq} = \emptyset$;
- (b) $I'(AT) = 1 - \frac{1}{(|U|+1)^2} \left(|U|^2(1 - I(AT)) + |[x_7]_{AT}^{\supseteq}| + |[x_7]_{AT}^{\subsetneq}| \right) = \frac{37}{49}$, likewise:
 $I'(AT - \{a_1\}) = \frac{36}{49}$, $I'(AT - \{a_2\}) = \frac{37}{49}$;
- (c) Since $I'(AT) > I'(AT - \{a_1\})$, we have the attribute core of the new table:
 $Core'(AT) = Core(AT) \cup \{a_1\} = \{a_1, a_3, a_4\}$.

Table 1 A SOIS about the ability of language

U	a_1	a_2	a_3	a_4
x_1	$\{E\}$	$\{E\}$	$\{F, G\}$	$\{F, G\}$
x_2	$\{E, F, G\}$	$\{E, F, G\}$	$\{F, G\}$	$\{E, F, G\}$
x_3	$\{E, G\}$	$\{E, F\}$	$\{F, G\}$	$\{F, G\}$
x_4	$\{E, F, G\}$	$\{E, F, G\}$	$\{E, G\}$	$\{E, F, G\}$
x_5	$\{E, F\}$	$\{F, G\}$	$\{E, F, G\}$	$\{E, G\}$
x_6	$\{E, F\}$	$\{E, G\}$	$\{F, G\}$	$\{E, F\}$

Table 2 The object inserted into the SOIS

U	a_1	a_2	a_3	a_4
x_7	$\{E, F, G\}$	$\{E, F\}$	$\{F\}$	$\{F, G\}$

- When the object x_6 is deleted from the Table 1, the attribute core is updated as follows:

- (a) $[x_6]_{AT}^{\supseteq} = \{x_2, x_6\}$; $[x_6]_{AT}^{\subsetneq} = \emptyset$; $[x_6]_{AT-\{a_1\}}^{\supseteq} = \{x_2, x_6\}$; $[x_6]_{AT-\{a_1\}}^{\subsetneq} = \emptyset$;
 $[x_6]_{AT-\{a_2\}}^{\supseteq} = \{x_2, x_6\}$; $[x_6]_{AT-\{a_2\}}^{\subsetneq} = \emptyset$; $[x_6]_{AT-\{a_3\}}^{\supseteq} = \{x_2, x_4, x_6\}$;
 $[x_6]_{AT-\{a_3\}}^{\subsetneq} = \emptyset$; $[x_6]_{AT-\{a_4\}}^{\supseteq} = \{x_2, x_6\}$; $[x_6]_{AT-\{a_4\}}^{\subsetneq} = \{x_1\}$;
- (b) $I'(AT) = 1 - \frac{1}{(|U|-1)^2} (|U|^2(1 - I(AT)) - |[x_6]_{AT}^{\supseteq}| - |[x_6]_{AT}^{\subsetneq}|) = \frac{17}{25}$, likewise:
 $I'(AT - \{a_3\}) = \frac{14}{25}$, $I'(AT - \{a_4\}) = \frac{17}{25}$;
- (c) Since $I'(AT) = I'(AT - \{a_4\})$, we have the attribute core of the new table:
 $Core'(AT) = Core(AT) - \{a_4\} = \{a_3\}$.

5 Conclusion

In this paper, we proposed incremental updating methods for computing the attribute core in SOIS based on information entropy. Through analyzing the changing mechanisms of information entropy when the information system is updated by inserting or deleting an object, we proved that there exists an inclusion relationship between attribute cores in the original information system and the updated ones. Then, we proposed the incremental methods to compute the core without re-implementing the static algorithm in a dynamic environment. The proposed approaches can reduce the computation time through updating the attribute core by partly modifying original attribute core instead of recalculation. Finally, an illustrated example was presented to verify the proposed methods.

Acknowledgments This work is supported by the National Science Foundation of China (Nos. 61175047, 61100117 and 71201133) and NSAF (No. U1230117), the Youth Social Science Foundation of the Chinese Education Commission (11YJC630127) and the Fundamental Research Funds for the Central Universities (SWJTU11ZT08, SWJTU12CX091, SWJTU12CX117).

References

1. Pawlak Z (1982) Rough sets. Int J Comput Inf Sci 11(5):341–356
2. Guan YY, Wang HK (2006) Set-valued information systems. Inf Sci 176:2507–2525
3. Qian YH, Dang CY, Liang JY, Tang DW (2009) Set-valued ordered information systems. Inf Sci 179(16):2809–2832

4. Bryson N, Mobolurin A (1997) Action learning evaluation procedure for multiple criteria decision making problems. *Eur J Oper Res* 96(2):379–386
5. Greco SB, Matarazzo B, Slowinski R (1999) Rough approximation of a preference relation by dominance relations. *Eur J Oper Res* 117(1):63–83
6. Greco S, Matarazzo B, Slowinski R (2007) Dominance-based Rough Set approach as a proper way of handling graduality in rough set theory. *Trans Rough Sets VII Lect Notes Comput Sci* 4400:36–52
7. Xu WH, Zhang XY, Zhang WX (2009) Knowledge granulation, knowledge entropy and knowledge uncertainty measure in ordered information systems. *Appl Soft Comput* 9:1244–1251
8. Fayyad UM, Shapiro LP, Smyth P, Uthurusamy R (1996) Advances in knowledge discovery and data mining. Park M (ed), AAAI Press/The MIT Press, California
9. Cercone V, Tsuchiya M (1993) Luesy editor's introduction. *IEEE Trans Knowl Data Eng* 5(6):901–902
10. Tong LY, An LP (2002) Incremental learning of decision rules based on rough set theory. *World Congress on Intelligent Control and Automation*, pp 420–425
11. Hu F, Dai J, Wang GY (2007) Incremental algorithms for attribute reduction in decision table. *Control Decis* 22(3):268–272
12. Yang M (2006) An incremental updating algorithm of the computation of a core based on the improved discernibility matrix. *Chin J Comput* 29(3):407–413 (in Chinese)
13. Wang GY, Yu H, Yang DC (2002) Decision table reduction based on conditional information entropy. *Chin J Comput* 25(7):759–766 (in Chinese)
14. Miao DQ, Hu GR (1999) A heuristic algorithm for reduction of knowledge. *Chin J Comput Res Dev* 36(6):681–684 (in Chinese)
15. Xu ZY, Yang BR, Guo YP (2006) Quick algorithm for computing core based on information entropy. *Mini-Micro Syst* 27:1711–1714 (in Chinese)
16. Fayyad UM, Shapiro LP, Smyth P, Uthurusamy R (1996) Advances in knowledge discovery and data mining. In: Park M (ed) AAAI Press/The MIT Press, California
17. Li TR, Ruan D, Geert W, Song J, Xu Y (2007) A rough sets based characteristic relation approach for dynamic attribute generalization in data mining. *Knowl Based Syst* 20(5):485–494
18. Chen HM, Li TR, Ruan D (2012) Maintenance of approximations in incomplete ordered decision systems while attribute values coarsening or refining. *Knowl Based Syst* 31:140–161
19. Zhang JB, Li TR, Ruan D, Liu D (2012) Rough sets based matrix approaches with dynamic attribute variation in set-valued information systems. *Int J Approximate Reasoning* 53(4):620–635
20. Yang M (2006) An incremental updating algorithm of the computation of a core based on the improved discernibility matrix. *Chin J Comput* 29(3):407–413 (in Chinese)
21. Jia XY, Shang L, Ji YS, Li WW (2007) An incremental updating algorithm for core computing in dominance-based rough set model. *Lect Notes Comput Sci*, 403–410
22. Liang JY, Wei W, Qian YH (2008) An incremental approach to computation of a core based on conditional entropy. *Syst Eng Theory Pract* 4(4):81–89 (in Chinese)

An Algebraic Analysis for Binary Intuitionistic L-Fuzzy Relations

Xiaodong Pan and Peng Xu

Abstract From the point of view of algebraic logic, this paper presents an algebraic analysis for binary intuitionistic lattice-valued fuzzy relations based on lattice implication algebras, which is a kind of lattice-valued propositional logical algebras. By defining suitable operations, we prove that the set of all binary intuitionistic lattice-valued fuzzy relations is a lattice-valued relation algebra, and some important properties are also obtained. This research shows that the algebraic description is advantageous to studying of structure of intuitionistic fuzzy relations.

Keywords Intuitionistic L-fuzzy relation · L-fuzzy relation · Lattice-valued relation algebra · Lattice implication algebra

1 Introduction

Fuzzy relations, introduced by Zadeh [24] in 1965, permit the gradual assessment of relativity between objects, and this is described with the aid of a membership function valued in the real unit interval $[0,1]$. Later on, Goguen [9] generalized this concept in 1967 to lattice-valued fuzzy relation (L-fuzzy relation for short) for an arbitrary complete Brouwerian lattice L instead of the unit interval $[0,1]$. Intuitionistic fuzzy relation, defined by Atanassov in 1984 [1–3], is another kind of generalization for Zadeh’s fuzzy relation, give us the possibility to model hesitation and uncertainty by using an additional degree. An intuitionistic L-fuzzy relation (ILFR for short) R between two universes U and V is defined as an intuitionistic L-fuzzy set (ILFS) in $U \times V$, assigns to each element $(x, y) \in U \times V$

X. Pan (✉) · P. Xu
School of Mathematics, Southwest Jiaotong University, Sichuan 610031 Chengdu,
P.R. China
e-mail: xdpan1@163.com

a membership degree $\mu_R(x, y) (\in L)$ and a non-membership degree $\nu_R(x, y) (\in L)$ such that $\mu_R(x, y) \leq N(\nu_R(x, y))$, where $N : L \rightarrow L$ is an involutive order-reversing operation on the lattice (L, \leq) . When $L = [0, 1]$, the object R is an intuition fuzzy relation (IFR) and the following condition holds: $(\forall (x, y) \in U \times V)(0 \leq \mu_R(x, y) + \nu_R(x, y) \leq 1)$. For all $(x, y) \in U \times V$, the number $\pi_R(x, y) = 1 - \mu_R(x, y) - \nu_R(x, y)$ is called the hesitation degree or the intuitionistic index of (x, y) to R . As we know, the structures and properties of operations on intuitionistic fuzzy relations have always been important topics in ILFS community. For that, P. Burillo and H. Bustince discussed the properties of several kinds of operations on ILFRs with different t-norms and t-conorms and characterized the structures of ILFRs in [4, 5]. In [6], G. Deschrijver and E. E. Kerre probed into the triangular compositions of ILFRs. In addition, the applications of ILFRs have also been developed rapidly in recent years, see [11, 13, 23].

As a fundamental conceptual and methodological tool in computer science just like logic, since the mid-1970s, relation algebras have been used intensively in applications of mathematics and computer science, and it also provides an apparatus for an algebraic analysis of ordinary predicate calculus, see e.g., [10, 18]. Following the idea of classical relation algebras, fuzzy relation algebras have been also considered by several researchers in [7, 8, 12]. Taking L to be a complete Heyting algebra, Furusawa developed the fuzzy relational calculus [7], and proved the representation theorems for algebraic formalizations of fuzzy relations in 1998. Furusawa's fuzzy relations algebras are equipped with sup-min composition and a semiscalar multiplication. In [8], Furusawa continued to study Dedekind categories with a cutoff operator, in which the Dedekind formula holds. In [16, 17], Popescu established the notion of MV-relation algebra based on MV-algebras, investigated their basic properties, and given a characterization of the "natural" MV-relation algebras. In order to provide a kind of algebraic model for lattice-valued first-order logic, based on complete lattice implication algebras [22], Pan introduced the notion of lattice-valued fuzzy relation algebra [14]; its basic properties and cylindric filters have also been established. For other categorical description of fuzzy relations, we refer readers to [19–21].

Along the line of the algebraic formalization of fuzzy relations, the present paper aims at investigating the arithmetical properties of mathematical structures formed by binary ILFRs based on the theory of L-fuzzy relation algebras (LRA). The paper is organized as follows. We recall some fundamental notions and properties of lattice implication algebras, Intuitionistic L-fuzzy relations and LRA in Sect. 2.2. Section 2.3 devotes to arithmetical properties of ILFRs. We conclude the paper in Sect. 2.4 with some radical suggestions for further problems.

2 Preliminaries

In this section, we review some basic definitions about intuitionistic L-fuzzy relations, lattice implication algebras, and lattice-valued relation algebras for the

purpose of reference and also recall some basic results which will be frequently used in the following, and we will not cite them every time they are used.

Definition 1 [22] A bounded lattice (L, \vee, \wedge, O, I) with order-reversing involution $'$ and a binary operation \rightarrow is called a lattice implication algebra if it satisfies the following axioms:

$$(I_1) \quad x \rightarrow (y \rightarrow z) = y \rightarrow (x \rightarrow z),$$

$$(I_2) \quad x \rightarrow x = I,$$

$$(I_3) \quad x \rightarrow y = y' \rightarrow x',$$

$$(I_4) \quad x \rightarrow y = y \rightarrow x = I \Rightarrow x = y,$$

$$(I_5) \quad (x \rightarrow y) \rightarrow y = (y \rightarrow x) \rightarrow x,$$

$$(L_1) \quad (x \vee y) \rightarrow z = (x \rightarrow z) \wedge (y \rightarrow z),$$

$$(L_2) \quad (x \wedge y) \rightarrow z = (x \rightarrow z) \vee (y \rightarrow z),$$

for all $x, y, z \in L$.

Example 1 [22] Let $L = [0, 1]$. If for any $a, b \in L$, put

$$a \vee b = \max\{a, b\}, a \wedge b = \min\{a, b\}, a' = 1 - a,$$

$$a \otimes b = \max\{0, a + b - 1\}, a \rightarrow b = \min\{1, 1 - a + b\}.$$

Then, $(L, \vee, \wedge, \otimes, \rightarrow, ')$ is called the Łukasiewicz algebra. Here, \rightarrow is called the Łukasiewicz implication, \otimes is called the Łukasiewicz product, and another operation \oplus , $a \oplus b = \min\{1, a + b\}$, is called the Łukasiewicz sum. It is easy to show that $(L, \vee, \wedge, \rightarrow, ', 0, 1)$ is also a lattice implication algebra.

Theorem 1 [22] Let $(L, \vee, \wedge, ', \rightarrow, 0, 1)$ be a lattice implication algebra, then (L, \vee, \wedge) is a distributive lattice.

In what follows, we list some well-known properties of lattice implication algebras that will often be used without mention. In a lattice implication algebra L , we define binary operation \otimes and \oplus as follows: for any $x, y \in L$, $a \otimes b = (a \rightarrow b)'$; $a \oplus b = a' \rightarrow b$.

Proposition 1 Let L be a lattice implication algebra, then for any $x, y, z \in L$, we have the following:

$$(1) \quad x \rightarrow y = I \text{ if and only if } x \leq y;$$

$$(2) \quad x \rightarrow y \geq x' \vee y;$$

$$(3) \quad (x \otimes y)' = x' \oplus y', (x \oplus y)' = x' \otimes y';$$

$$(4) \quad O \otimes x = O, I \otimes x = x, x \otimes x' = O, O \oplus x = x, I \oplus x = I, x \oplus x' = I;$$

$$(5) \quad x \otimes y = (x \vee y) \otimes (x \wedge y), x \oplus y = (x \vee y) \oplus (x \wedge y);$$

$$(6) \quad x \otimes (y \vee z) = (x \otimes y) \vee (x \otimes z), x \otimes (y \wedge z) = (x \otimes y) \wedge (x \otimes z).$$

In [14], by generalizing the classical (Boolean) relation algebras, we introduce the notion of lattice-valued relation algebra based on complete lattice implication algebras L .

Definition 2 A lattice-valued relation algebra (**LRA** for short) is an algebra $\mathfrak{L} = (L, \vee, \wedge, ', \rightarrow, O, I, ;, \smile, \Delta)$, where

- (1) $(L, \vee, \wedge, ', \rightarrow, O, I)$ is a lattice implication algebra;
- (2) $(L, ;, \Delta)$ is a monoid (a semigroup with identity, Δ);
- (3) $(x \vee y); z = (x; z) \vee (y; z)$;
- (4) $(x \rightarrow y)^\smile = x^\smile \rightarrow y^\smile$;
- (5) $x^{\smile\smile} = x$;
- (6) $\Delta' \vee \Delta = I$;
- (7) $(x; y)^\smile = y^\smile; x^\smile$;
- (8) $y' = y' \vee (x^\smile; (x; y)')$;
- (9) $(a \otimes x); (b \otimes y) \leq (a; b) \otimes (x; y)$.

From the above definition, an **LRA** is an expansion of the corresponding lattice implication algebra with the operations $\smile, ;, \Delta$, where \smile is called the converse operation and $;$ the relative multiplication operation and Δ the diagonal element. The class of **LRAs** will be denoted by \mathfrak{LRA} .

Let $(L, \vee, \wedge, ', \rightarrow, 0, 1)$ be a complete lattice implication algebra. An intuitionistic L-fuzzy relation R from a universe U to a universe V is an intuitionistic fuzzy set in $U \times V$, i.e. an object having the form $R = \{((x, y), \mu_R(x, y), \nu_R(x, y)) | x \in U, y \in V\}$, where $\mu_R : U \times V \rightarrow L$ and $\nu_R : U \times V \rightarrow L$ satisfy the condition $(\forall (x, y) \in U \times V)(\mu_R(x, y) \leq (\nu_R(x, y))')$. The set of all intuitionistic L-fuzzy relations from U to V will be denoted by $\mathfrak{IFR}(U \times V)$. We say the intuitionistic L-fuzzy relation R is contained in the intuitionistic L-fuzzy relation S , written $R \subseteq S$ if $\mu_R(x, y) \leq \mu_S(x, y)$ and $\nu_R(x, y) \geq \nu_S(x, y)$ for all $(x, y) \in U \times V$. The zero relation $O_{U \times V}$ and the full relation $I_{U \times V}$ are intuitionistic L-fuzzy relations with $\mu_{O_{U \times V}}(x, y) = O$, $\nu_{O_{U \times V}}(x, y) = I$ and $\mu_{I_{U \times V}}(x, y) = I$, $\nu_{I_{U \times V}}(x, y) = O$ for all $(x, y) \in U \times V$, respectively. It is trivial that \subseteq is a partial order on $\mathfrak{IFR}(U \times V)$ and $O_{U \times V} \subseteq R \subseteq I_{U \times V}$ for all intuitionistic L-fuzzy relations R . The union, intersection complement, and converse of intuitionistic L-fuzzy relations are defined as follows:

$$\begin{aligned} R \cup S &= \{((x, y), \mu_R(x, y) \vee \mu_S(x, y), \nu_R(x, y) \wedge \nu_S(x, y)) | (x, y) \in U \times V\}; \\ R \cap S &= \{((x, y), \mu_R(x, y) \wedge \mu_S(x, y), \nu_R(x, y) \vee \nu_S(x, y)) | (x, y) \in U \times V\}; \\ R' &= \{((x, y), \nu_R(x, y), \mu_R(x, y)) | (x, y) \in U \times V\}; \\ R^\smile &= \{((x, y), \mu_R(y, x), \nu_R(y, x)) | (x, y) \in U \times V\}. \end{aligned}$$

It is easy to show that $(\mathfrak{IFR}(U \times V), \cup, \cap, O_{U \times V}, I_{U \times V})$ is a complete lattice. Let $R \in \mathfrak{IFR}(U \times V)$ and $S \in \mathfrak{IFR}(V \times W)$, the relative multiplication (or composition) $R; S$ is defined as follows:

$$R; S = \left((x, z), \bigvee_{y \in V} (\mu_R(x, y) \otimes \mu_S(y, z)), \bigwedge_{y \in V} (\nu_R(x, y) \oplus \nu_S(y, z)) | (x, z) \in U \times W \right),$$

then $\bigvee_{y \in V} (\mu_R(x, y) \otimes \mu_S(y, z)) \leq (\bigwedge_{y \in V} (v_R(x, y) \oplus v_S(y, z)))'$ for any $(x, z) \in U \times W$. Let $R, S \in \mathfrak{IFLFR}(U \times V)$, $R \rightarrow S$ is also an intuitionistic L-fuzzy relation on $U \times V$, where $\mu_{R \rightarrow S}(x, y)$ means the truth degree of the sentence, if xRy , then xSy ; and $v_{R \rightarrow S}(x, y)$ means the truth degree of the sentence, if xRy does not hold, then xSy does not hold.

Throughout this paper, unless otherwise stated, L always denotes a complete lattice implication algebra. For more details of lattice implication algebras, we refer readers to [15, 22].

3 The Arithmetical Properties of $\mathfrak{IFLFR}(U \times U)$

In this section, we study the arithmetical properties of intuitionistic L-fuzzy relations on the universe U .

Theorem 2 *Let U be any non-empty set. The union, intersection, complement, converse, and composition of intuitionistic L-fuzzy relations in U are defined as that in Sect. 2.2. We also define the operations $\rightarrow, \oplus, \otimes$ on $\mathfrak{IFLFR}(U \times U)$, for any $R, S \in \mathfrak{IFLFR}(U \times U)$,*

$$\begin{aligned} R \rightarrow S &= ((x, y), v_R(x, y) \oplus \mu_S(x, y), \mu_R(x, y) \otimes v_S(x, y)) | (x, y) \in U \times U, \\ R \oplus S &= ((x, y), \mu_R(x, y) \oplus \mu_S(x, y), v_R(x, y) \otimes v_S(x, y)) | (x, y) \in U \times U, \\ R \otimes S &= ((x, y), \mu_R(x, y) \otimes \mu_S(x, y), v_R(x, y) \oplus v_S(x, y)) | (x, y) \in U \times U, \end{aligned}$$

and define the identity relation Δ_U is an intuitionistic L-fuzzy relation on U such that for any $(x, y) \in U \times U$,

$$\mu_{\Delta_U}(x, y) = \begin{cases} I & \text{if } x = y, \\ O, & \text{otherwise.} \end{cases} \quad \text{and} \quad v_{\Delta_U}(x, y) = \begin{cases} O & \text{if } x = y, \\ I, & \text{otherwise.} \end{cases}$$

Then, $(\mathfrak{IFLFR}(U \times U), \cup, \cap, ', \rightarrow, O_U, I_U, ;, \smile, \Delta_U)$ is an **LRA**.

Proof The zero relation and the full relation on U are denoted by O_U and I_U , respectively. As we have discussed above, $(\mathfrak{IFLFR}(U \times U), \cup, \cap, O_U, I_U)$ is a bounded lattice. For any $R, S, T \in \mathfrak{IFLFR}(U \times U)$, since

$$\begin{aligned} \mu_{R \rightarrow (S \rightarrow T)}(x, y) &= v_R(x, y) \oplus \mu_{S \rightarrow T}(x, y) = v_R(x, y) \oplus (v_S(x, y) \oplus \mu_T(x, y)) \\ &= v_S(x, y) \oplus (v_R(x, y) \oplus \mu_T(x, y)) = v_S(x, y) \oplus \mu_{R \rightarrow T}(x, y) \\ &= \mu_{S \rightarrow (R \rightarrow T)}(x, y), \end{aligned}$$

and $v_{R \rightarrow (S \rightarrow T)}(x, y) = v_{S \rightarrow (R \rightarrow T)}(x, y)$ can also be proved similarly, I_1 holds. In this way, we can validate I_2, \dots, I_5 and L_1, L_2 . Hence, $(\mathfrak{IFLFR}(U \times U), \cup, \cap, ', \rightarrow, O_U, I_U)$ is a lattice implication algebra. The associativity $(R; S); T = R; (S; T)$ and the unitary law $R; \Delta_U = \Delta_U; R = R$ and the zero law $R; O_U = O_U; R = O_U$ are obvious. So far, we have proved (1), (2) in definition 2. (4), (5), and (6) are

obvious. The rest is to prove (3), (7), (8), and (9). Note that L is an infinite distributive lattice, hence,

$$\begin{aligned}
v_{(R \cup S);T}(x, z) &= \bigwedge_{y \in U} ((v_R(x, y) \wedge v_S(x, y)) \oplus v_T(y, z)) \\
&= \bigwedge_{y \in U} ((v_R(x, y) \oplus v_T(y, z)) \wedge (v_S(x, y) \oplus v_T(y, z))) \\
&= \bigwedge_{y \in U} ((v_R(x, y) \oplus v_T(y, z)) \wedge \bigwedge_{y \in U} ((v_S(x, y) \oplus v_T(y, z)))) \\
&= v_{R;T}(x, z) \wedge v_{S;T}(x, z) = v_{(R;T) \cup (S;T)}(x, z).
\end{aligned}$$

$\mu_{(R \cup S);T}(x, z) = \mu_{(R;T) \cup (S;T)}(x, z)$ is similar. Thus, (3) holds. In this way, we can obtain (7). In order to prove (8), we only need to prove $R^\sim; (R; S)' \leq S'$, or $(R^\sim; (R; S)')' \geq S$. As

$$\begin{aligned}
\mu_{(R^\sim; (R; S)')'}(x, y) &= \bigwedge_{z \in U} ((\mu_{R^\sim}(x, z))' \oplus \bigvee_{w \in U} (\mu_R(z, w) \otimes \mu_S(w, y))) \\
&= \bigwedge_{z \in U} ((\mu_R(z, x))' \oplus \bigvee_{w \in U} (\mu_R(z, w) \otimes \mu_S(w, y))) \\
&= \bigwedge_{z \in U} \bigvee_{w \in U} ((\mu_R(z, x))' \oplus (\mu_R(z, w) \otimes \mu_S(w, y)));
\end{aligned}$$

and for any $z \in U$,

$$\begin{aligned}
&\bigvee_{w \in U} ((\mu_R(z, x))' \oplus (\mu_R(z, w) \otimes \mu_S(w, y))) \\
&= \bigvee_{w \in U, w \neq x} ((\mu_R(z, x))' \oplus (\mu_R(z, w) \otimes \mu_S(w, y))) \bigvee \\
&\quad ((\mu_R(z, x))' \oplus (\mu_R(z, x) \otimes \mu_S(x, y))) \\
&= \bigvee_{w \in U, w \neq x} ((\mu_R(z, x))' \oplus (\mu_R(z, w) \otimes \mu_S(w, y))) \bigvee \\
&\quad ((\mu_R(z, x) \rightarrow (\mu_R(z, x) \rightarrow (\mu_S(x, y))')')') \\
&= \bigvee_{w \in U, w \neq x} ((\mu_R(z, x))' \oplus (\mu_R(z, w) \otimes \mu_S(w, y))) \bigvee \\
&\quad ((\mu_S(x, y) \rightarrow (\mu_R(z, x))') \rightarrow (\mu_R(z, x))') \\
&= \bigvee_{w \in U, w \neq x} ((\mu_R(z, x))' \oplus (\mu_R(z, w) \otimes \mu_S(w, y))) \bigvee (\mu_S(x, y) \vee (\mu_R(z, x))') \\
&\geq \mu_S(x, y).
\end{aligned}$$

Similarly, we can prove $v_{(R^\sim; (R; S)')'}(x, y) \leq v_S(x, y)$. Thus, (8) holds. For (9), we only validate the part of the degree of membership, the other part is similar.

$$\begin{aligned}
\mu_{(S \otimes R);(T \otimes U)}(x, y) &= \bigvee_{z \in U} \left(\mu_{S \otimes R}(x, z) \otimes \mu_{T \otimes U}(z, y) \right) \\
&= \bigvee_{z \in U} \left(\mu_S(x, z) \otimes \mu_R(x, z) \otimes \mu_T(z, y) \otimes \mu_U(z, y) \right) \\
&= \bigvee_{z \in U} \left(\mu_S(x, z) \otimes \mu_T(z, y) \otimes \mu_R(x, z) \otimes \mu_U(z, y) \right) \\
&\leq \bigvee_{z \in U} \left(\mu_S(x, z) \otimes \mu_T(z, y) \right) \otimes \bigvee_{z \in U} \left(\mu_R(x, z) \otimes \mu_U(z, y) \right) \\
&= \mu_{(S;T) \otimes (R;U)}(x, y).
\end{aligned}$$

Thus, (9) holds. To sum up, $(\mathfrak{I}\mathfrak{L}\mathfrak{F}\mathfrak{R}(U \times U), \cup, \cap, ', \rightarrow, O_U, I_U, ;, \smile, \Delta_U)$ is an **LRA**.

Corollary 1 *Let U be any non-empty set. For any $R, S \in \mathfrak{I}\mathfrak{L}\mathfrak{F}\mathfrak{R}(U \times U)$, the following identities are true:*

- (1) $I_U^\smile = I_U, \Delta_U^\smile = \Delta_U, O_U^\smile = O_U$;
- (2) $(R')^\smile = (R^\smile)'$;
- (3) $(R \vee S)^\smile = R^\smile \vee S^\smile, (R \wedge S)^\smile = R^\smile \wedge S^\smile$;
- (4) $(R \oplus S)^\smile = R^\smile \oplus S^\smile, (R \otimes S)^\smile = R^\smile \otimes S^\smile$.

According to Definition 2, the following conclusions can be obtained. Here, we only prove (5); for more details, please refer to [14].

Proposition 2 *Let U be any non-empty set. For any $R, S, T \in \mathfrak{I}\mathfrak{L}\mathfrak{F}\mathfrak{R}(U \times U)$, the following conclusions are true:*

- (1) $R; (S \vee T) = (R; S) \vee (R; T)$;
- (2) $S' = S' \vee ((S; R^\smile)'; R)$;
- (3) $R; S \leq T' \Leftrightarrow R^\smile; T \leq S' \Leftrightarrow T; S^\smile \leq R'$;
- (4) $R \leq I_U; R; I_U, I_U = I_U; I_U$;
- (5) $R; (S \oplus T) \leq (R; S) \oplus (I_U; T), (R \oplus S); T \leq (R; T) \oplus (S; I_U)$;
- (6) $I_U; (R \rightarrow S); I_U \leq (I_U; R'; I_U)' \rightarrow (I_U; S; I_U)$;
- (7) $(I_U; R'; I_U)' \leq I_U; R; I_U$;
- (8) $R^\smile \leq S^\smile$ if and only if $R \leq S$.

Proof (5) We only prove the first inequality, the second can be proved similarly. Since

$$\begin{aligned}
(R^\smile; (R; S)') \otimes (I_U; (I_U; T)') &\geq (R^\smile \otimes I_U); ((R; S)' \otimes (I_U; T)') \\
&= R^\smile; ((R; S)' \otimes (I_U; T)'),
\end{aligned}$$

and note that $R^\smile; (R; S)' \leq S'$ and $I_U; (I_U; T)' = I_U^\smile; (I_U; T) \leq T'$, thus

$$R^\smile; ((R; S)' \otimes (I_U; T)') \leq S' \otimes T' = (S \oplus T)',$$

this is equivalent to $R^\sim; (S \oplus T) \leq ((R; S)' \otimes (I_U; T)')' = (R; S) \oplus (I_U; T)$.

Corollary 2 *Let U be any non-empty set. For any $R, S \in \mathfrak{ILFR}(U \times U)$, then,*

- (1) $R = R^\sim$, or R and R^\sim are incomparable, in notation, $R \parallel R^\sim$;
- (2) if $S \leq \Delta_U$, then $R \geq R; S$ and $R \geq S; R$; if $S \geq \Delta_U$, then $R \leq R; S$ and $R \leq S; R$.

In the following, for any $R, S \in \mathfrak{ILFR}(U \times U)$, we let $\square_U = \Delta'_U$, $R \dagger S = (R'; S')'$, where \square_U is called the diversity element in $\mathfrak{ILFR}(U \times U)$, \dagger is called relative addition. By Theorem 2, Corollary 1, and Proposition 2, the following conclusions can be obtained, the proofs are omitted.

Proposition 3 *Let U be any non-empty set. For any $R, S, T \in \mathfrak{ILFR}(U \times U)$, then,*

- (1) $R \dagger (S \dagger T) = (R \dagger S) \dagger T$;
- (2) $(R \dagger S)^\sim = S^\sim \dagger R^\sim$;
- (3) $R^\sim; (R' \dagger S) \leq S$;
- (4) $\square_U^\sim = \square_U$;
- (5) $R \dagger (S \wedge T) = (R \dagger S) \wedge (R \dagger T)$, $(R \wedge S) \dagger T = (R \dagger T) \wedge (S \dagger T)$;
- (6) $R; (S \dagger T) \leq (R; S) \dagger T$;
- (7) $(R \dagger S); T \leq R \dagger (S; T)$;
- (8) $\Delta_U \leq R \dagger (R')^\sim$, $R; (R')^\sim \leq \square_U$;
- (9) $\Delta_U \leq R \dagger S$ if and only if $\Delta_U \leq S \dagger R$ if and only if $\Delta_U \leq R^\sim \dagger S^\sim$ if and only if $\Delta_U \leq S^\sim \dagger R^\sim$;
- (10) $R \dagger \square_U = \square_U \dagger R = R$, $R \dagger I_U = I_U \dagger R = I_U$;

Theorem 3 *Let U be any non-empty set and $R, S, P \in \mathfrak{ILFR}(U \times U)$. Then, $(S^\sim; R')' = \bigvee_{S; T \leq R} T$. In particular, $P^\sim = \bigvee_{P'; T \leq \square_U} T$.*

Proof For the first equality, on the one hand, by Definition 2, we have $S; (S^\sim; R')' \leq R$; on the other hand, if $S; T \leq R$, by Proposition 2, this is equivalent to $S^\sim; R' \leq T'$, that is $T \leq (S^\sim; R')'$. Hence, $(S^\sim; R')' = \bigvee_{S; T \leq R} T$. Let $S = P'$, $R = \square_U$, we can obtain the second equality.

In what follows, we prove the main properties of the ILFRs using the calculus of the algebraic theory. Let $R \in \mathfrak{ILFR}(U \times U)$. We say that R is reflexive if $R \geq \Delta_U$, antireflexive if $R \leq \square_U$, symmetrical if $R = R^\sim$, and transitive if $R \geq R; R$.

Theorem 4 *Let U be any non-empty set and $R, S \in \mathfrak{ILFR}(U \times U)$. Then,*

- (1) R is reflexive if and only if R' is antireflexive;
- (2) R is symmetrical if and only if R' is symmetrical;
- (3) R is transitive if and only if R^\sim is transitive.

Proof

- (1) R is reflexive if and only if $R \geq \Delta_U$ if and only if $R' \leq (\Delta_U) = \square_U$ if and only if R' is antireflexive.
- (2) By Corollary III.1 (2), we have $(R')^\sim = (R^\sim)'$, so R is symmetrical if and only if $R = R^\sim$ if and only if $R' = (R^\sim)' = (R')^\sim$ if and only if R' is symmetrical.
- (3) If R is transitive, then $R \geq R;R$, by Proposition III.1(8), we have $R^\sim \geq (R;R)^\sim = R^\sim;R^\sim$; hence, R^\sim is transitive, and vice versa.

Corollary 3 *Let U be any non-empty set and $R \in \mathfrak{I}\mathfrak{Q}\mathfrak{F}\mathfrak{R}(U \times U)$. The following conclusions hold:*

- (1) *if R is reflexive, then so are $R;R$ and R^\sim ;*
- (2) *if R is symmetrical, then so are $R;R$ and R^\sim ;*
- (3) *if R is transitive, then so are $R;R$ and R^\sim ;*
- (4) *if R is antireflexive, then $R \geq R \dagger R$.*

4 Conclusion

In this paper, we present an algebraic analysis for binary intuitionistic lattice-valued fuzzy relations based on lattice implication algebras. For the theory and application of intuitionistic L-fuzzy relations, the algebraic description shows its advantages. More importantly, by the algebraization of the set of intuitionistic L-fuzzy relations, we can obtain a denotational semantics of intuitionistic L-fuzzy theory and hence a mathematical theory to reason about notions like correctness. Consequently, one may prove such properties using the calculus of the algebraic theory, the results, and methods of applications of fuzzy theory may be described by simple terms in this language.

Acknowledgments The work was partially supported by the National Natural Science Foundation of China (Grant No. 61100046, 61175055) and the application fundamental research plan project of Sichuan Province (Grant No. 2011JY0092), and the Fundamental Research Funds for the Central Universities (Grant No. SWJTU12CX054, SWJTU12ZT14).

References

1. Atanassov K, Stoeva S (1984) Intuitionistic L-fuzzy sets. In: Trappl R (ed) Cybernetics and Systems Research 2. Elsevier, Amsterdam, pp 539–540
2. Atanassov K (1984) Intuitionistic fuzzy relations. In: Antonov L (ed) Proceedings of the Third international symposium on automation and scientific instrumentation. Varna, vol II:56–57

3. Atanassov K (1986) Intuitionistic fuzzy sets. *Fuzzy Sets Syst* 20:87–96
4. Burillo P, Bustince H (1995) Intuitionistic fuzzy relations (Part I). *Mathware Soft Comput* 2:5–38
5. Burillo P, Bustince H (1995) Intuitionistic fuzzy relations (Part II). *Mathware Soft Comput* 2:117–148
6. Deschrijver G, Kerre EE (2003) On the composition of intuitionistic fuzzy relations. *Fuzzy Sets Syst* 136:333–361
7. Furusawa H (1998) Algebraic Formalisations of Fuzzy Relations and their Representation Theorems. Kyushu University, PhD-Thesis
8. Furusawa H, Kawahara Y, Winter M (2011) Dedekind Categories with cutoff operators. *Fuzzy Sets Syst* 173:1–24
9. Goguen JA (1967) L-fuzzy sets. *J Math Anal Appl* 18:145–157
10. Halmos P, Givant SR (1998) *Logic as algebra*. The Mathematical Association of America, Washington, D.C.
11. Hwang CM, Yang MS, Hung WL et al (2012) A similarity measure of intuitionistic fuzzy sets based on the Sugeno integral with its application to pattern recognition. *Inf Sci* 189:93–109
12. Kawahara Y, Furusawa H (1999) An algebraic formalization of fuzzy relations. *Fuzzy Sets Syst* 101:125–135
13. Li DF (2005) Multiattribute decision making models and methods using intuitionistic fuzzy sets. *J Comput Syst Sci* 70:73–85
14. Pan XD, Xu Y (2013) On the algebraic structure of binary lattice-valued fuzzy relations. *Soft Computing* 17:411–420
15. Pan XD, Xu Y (2010) Semantic theory of finite lattice-valued propositional logic. *sci china. Inf Sci* 53:2022–2031
16. Popescu A (2005) Many-valued relation algebras. *Algebra Univers* 53:73–108
17. Popescu A (2007) Some algebraic theory for many-valued relation algebras. *Algebra Univers* 56:211–235
18. Tarski A (1941) On the calculus of relations. *J Symbolic Logic* 6:73–89
19. Winter M (2001) A new algebraic approach to L-fuzzy relations convenient to study crispness. *Inf Sci* 139:233–252
20. Winter M (2003) Representation Theory of Goguen Categories. *Fuzzy Sets Syst* 1:339–357
21. Winter M (2003) Goguen categories. *J Relat Methods Comput Sci* 138:85–126
22. Xu Y, Ruan D, Qin KY, Liu J (2003) *Lattice-valued logic-an alternative approach to treat fuzziness and incomparability*. Springer, Berlin
23. Xu ZS, Yager RR (2006) Some geometric aggregation operators based on intuitionistic fuzzy sets. *Int J Gen Syst* 35:417–433
24. Zadeh LA (1965) Fuzzy sets. *Inf. Control* 8:338–353

An Improved Classification Method Based on Random Forest for Medical Image

Niu Zhang, Haiwei Pan, Qilong Han and Xiaoning Feng

Abstract Medical image classification is an important part in domain-specific application image mining because there are several technical aspects which make this problem challenging. Ensemble methodologies have evolved to leverage potentially thousands of base classifiers that are usually instantiations of the same underlying model (e.g., neural networks, decision trees). Random forest (RF) is a classical ensemble classification algorithm which gives the bound of generalization error (or probability of misclassification) of ensemble classifier, but the bound cannot directly address application spaces in which error costs are inherently unequal. However, the error costs in medical image classification are inherently unequal. In this paper, we propose an improved classification algorithm based on random forest (IRFA) to solve the classification problem. It leverages key elements of the derivation of generalization error bound to derive bounds on detection rate (DET) and false alarms rate (FAR) on ROC and gives the performance optimization guidelines for tuning class-specific correlation inferred from the bounds for each region. At last, we use IRFA for medical image classification.

Keywords Medical image · Image mining · Domain knowledge · Ensemble classification · Random forest

1 Introduction

Advances in image acquisition and storage technology have led to tremendous growth in large and detailed image databases [1]. Although there are many relatively mature theories and techniques in image mining area, the image mining is

N. Zhang · H. Pan (✉) · Q. Han · X. Feng
College of Computer Science and Technology, Harbin Engineering University,
Harbin, Heilongjiang, China
e-mail: heaven_007cn@yahoo.com.cn

still in its infancy and gradually attended by experts and researchers. Medical image contains a wealth of hidden information that is useful for physicians make correct decision for a patient. Only through color, texture, shape, and other general characteristics for image feature extraction cannot be fully expressed the medical image. As medical domain knowledge is valuable experience accumulated via clinical diagnosis, it is important to use the domain knowledge for feature extraction in medical image classification.

Classification is a primary task of inductive learning in data mining and machine learning [2]. Many effective inductive learning techniques have developed such as naive Bayes, decision trees, neural networks, and ensemble classifier, and most classification algorithms assume the different types of misclassification have equal error cost. However, in many real-world applications, this assumption is not true and the differences between different misclassification errors can be quite large. For example, in medical diagnosis, missing a cancer diagnosis (false negative) is much more serious than the other way around (false positive); the patient could lose his/her life because of the delay in treatment. Cost-sensitive learning has received much attention in recent years to deal with such an issue [3]. So, it is very important to take account unequal cost regimes into medical image classification.

Random forest (RF) [4] is a widely used ensemble classification algorithm and gives the bound of the generalization error above by a function of the mean correlation between the constituent (i.e., base) classifiers and their average strength. This bound suggests that increasing the strength and/or decreasing the correlation of an ensemble's base classifiers may yield improved performance under the assumption of equal error costs. But its existing bounds do not directly address application spaces in which error costs are inherently unequal.

In this paper, we leverage key elements of Breiman's derivation of generalization error bound to derive bounds on detection rate (DET) and false alarms rate (FAR) on ROC and give the performance optimization guidelines for tuning class-specific correlation inferred from the bounds for each region.

The remaining of the paper is organized as follows. [Section 2](#) introduces the preprocessing of medical images. [Section 3](#) gives the details of IFRA. [Section 4](#) gives the experiment results and analysis, while the conclusion is given in [Sect. 5](#).

2 Medical Image Preprocessing

The prime objective of the preprocessing is to improve the image data quality by suppressing undesired distortions or enhancing the required image features for further processing. The irrelevant data present in the image has been eliminated using the preprocessing technique.

We use the CT images of brain as our research object. First, we use the adaptive water immersion algorithm [5] to extract the ROI (region of interest) from the CT images of brain (as shown in [Fig. 1](#)). Then, we extract relevant information as the

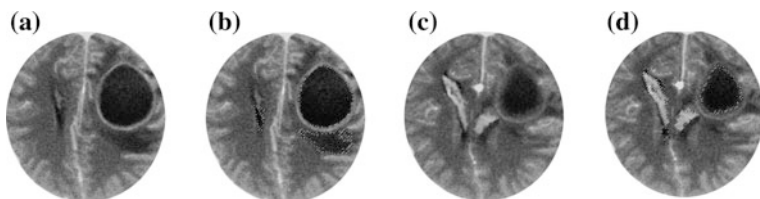


Fig. 1 **a** and **c** are the two original images of the brain, the *dotted lines* of **b** and **d** are the ROI marked by using the adaptive water immersion algorithm

Table 1 Feature database of medical image

IM ID	P_1	P_2	...	P_m	Class ID
IM ₁	1	1	...	0	Normal
IM ₂	0	1	...	0	Abnormal
...
IM _n	1	0	...	2	Normal

features for each ROI, such as the symmetry, area, location, elongation, and circle-like. Then, we clustered these ROI using DBSCAN clustering algorithm [6]. According to the clustering results, we count the number for each category ROI for each medical image. At last, according to the count for each medical image, we give the form of medical image feature expression (as shown in Table 1).

In Table 1, IM_{*i*} is the *i*th medical image, P_{*i*} is the *i*th cluster of the ROI, normal and abnormal are the two class labels of the medical image.

3 Improved Classification Algorithm Based on Random Forest

In this section, we first introduce the evaluation of RF, and then we give the ROC curve performance regions and the bounds of ROC. At last, we give the procession of IRFA.

3.1 The Evaluation of Random Forest

Ensemble methodologies have proven to be highly successful at reducing the generalization error in classification [7]. Placing a bound on the generalization error is beneficial both for characterizing the performance of ensemble classifiers in the field and in motivating efforts at ensemble classifier optimization. Generalization error bounds have previously been derived for ensemble classification methods by Refs. [4, 8, 9].

The bound derived by Breiman [3] is of particular interest. He demonstrated that as the number of base classifiers in the ensemble increases, the generalization error, E , converges and is bounded as follows:

$$E \leq \frac{\bar{\rho}(1 - s^2)}{s^2} \quad (1)$$

where $\bar{\rho}$ denotes the mean correlation of base classifier predictions, and s represents the average strength of the base classifiers. It is immediately apparent that the bound on generalization error decreases as the base classifiers become stronger and/or less correlated. However, it does not explicitly characterize the impact of the strength and correlation of base classifiers on class-specific error rates. To this end, we have developed extensions to Breiman's bound that directly address these error rates.

3.2 The ROC Curve Performance Regions

Vote frequencies generated by the ensemble are used to classify a data sample. When the positive and negative classes are associated with the labels 1 and 0, respectively, these votes can be combined to compute a numerical score, given as follows:

$$\text{score}(x) = \frac{2}{k} \sum_{k=1}^K h_k(x) - 1 \quad (2)$$

where k equals the number of base classifiers in the ensemble, and $h_k(x)$ is the label assigned by the k th base classifier to the input vector x . The score lies within the interval $[-1, 1]$ and relates directly to the margin function.

Based on Breiman's [4] margin function for an ensemble classifier, we give the margin function for an ensemble classifier for the two-class case as follows:

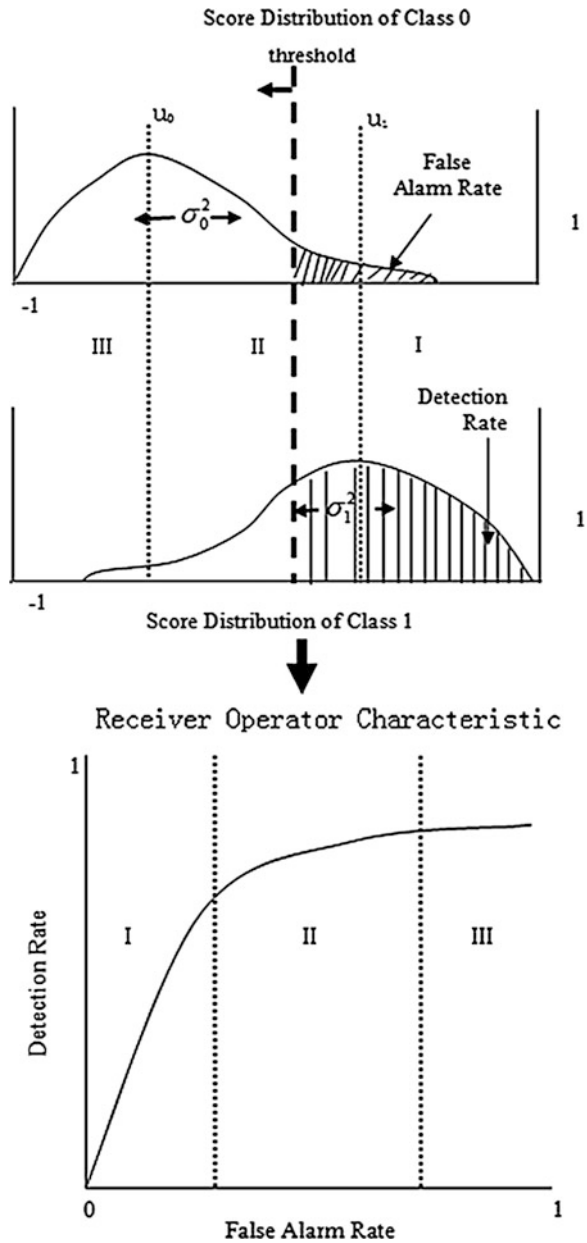
$$\text{mg}(x, y) = \frac{2}{k} \sum_{k=1}^K I(h_k(x) = y) - 1 \quad (3)$$

where $I(\cdot)$ is an indicator function, and y is the true class label associated with data sample x . The margin function measures the degree to which the votes for the correct class exceed the votes for the incorrect class; in essence, it is a measure of confidence.

It can be easily shown that for class 1 samples, the score is equal to the margin, and for class 0 samples, the score is the negative of the margin.

The scores computed for each class form distributions that can be used to generate a ROC curve. Each point on the ROC curve indicates the false alarm and detection rates of the ensemble classifier, given a fixed decision threshold. Consequently, the curve can be generated by sweeping a decision threshold across the

Fig. 2 Performance regions of a ROC curve



two class-specific score distributions simultaneously (as shown in Fig. 2). The probability mass to the right of this threshold for the positive and negative class score distributions corresponds to the detection and false alarm rates, respectively.

Breiman defines the average strength of the base classifiers as the expected value of the margin function. Leveraging the relationship between the score distributions and the margin function, we can estimate the class-specific strengths, s_0 and s_1 , by

$$\begin{aligned} s_0 &= -u_0 \\ s_1 &= u_1 \end{aligned} \tag{4}$$

where u_i is the mean of the score distribution for class i .

The overall strength s is a weighted average of the class-specific strengths, and it measures the degree of separation between the means of the score distributions. It can be written in terms of the class-specific strengths as follows:

$$s = \frac{n_1 s_1 + n_0 s_0}{n_1 + n_0} \tag{5}$$

where n_i is the number of class i samples.

The variance σ^2 of the margin function is also related to the strength and correlation of the base classifiers and can be expressed in general by the following inequality:

$$\sigma^2 \leq \bar{\rho}(1 - s^2) \tag{6}$$

We can write Eq. (6) in terms of the positive and negative classes as follows:

$$\begin{aligned} \sigma_0^2 &\leq \bar{\rho}_0(1 - s_0^2) \\ \sigma_1^2 &\leq \bar{\rho}_1(1 - s_1^2) \end{aligned} \tag{7}$$

where σ_i^2 is the variance of the class i score distribution and denotes the mean correlation between the base classifiers calculated for the class i samples.

From Eq. (7), we know that for fixed class-specific strength, reducing (or increasing) the class-specific correlation between the base classifiers can yield a corresponding shift in the variance of the margin function (and hence, the variance of the score distribution) for that class. We will discuss this relationship further in [Sect. 3.3](#).

3.3 The Bounds of ROC

Breiman gives the derivation of the generalization error under the assumption of equal error costs, not takes the specific class into consider. The decision threshold is implicitly fixed, so that the bound that derived is on a single point on the ROC curve. In this section, we will extend this bound to the entire ROC curve at different error cost and every decision threshold value must be considered.

When constructing the ROC curve, the false alarm rate (FAR) is the probability that a score exceeds some threshold t from the class 0 empirical score distribution. Similarly, the detection rate (DET) is the probability that a score exceeds t from the class 1 empirical score distribution. These rates can be expressed as follows:

$$\begin{aligned} \text{FAR} &= P(Z_0 \geq t) \\ \text{DET} &= P(Z_1 \geq t) \end{aligned} \quad (8)$$

where Z_0 and Z_1 are random variables representing the class-specific scores for a particular sample.

We can use the one-tailed Chebyshev inequality enables us to derive bounds on the false alarm rate and detection rate in terms of the class-specific strengths and correlations for a given threshold t . The one-tailed Chebyshev inequality as follows:

$$p(Z - u \geq k) \leq \frac{1}{1 + \frac{k^2}{\sigma^2}}, k > 0 \quad (9)$$

where u and σ^2 are the mean and variance of Z , Z is infinitely close to u .

For $t = k + u$, the Eq. (9) can be expressed as follows:

$$p(Z \geq t) \leq \frac{1}{1 + \frac{(t-u)^2}{\sigma^2}}, t > u \quad (10)$$

Equation (10) gives the one tail of score distribution where $t > u$. The other tail of this distribution can give by subtract both sides of the inequality from 1 to yield an inequality describing the region and can be expressed as follows:

$$p(Z \geq t) \geq \frac{1}{1 + \frac{\sigma^2}{(t-u)^2}}, t < u \quad (11)$$

Equations (10) and (11) give us two limits on the probability that a random variable Z will be greater than the threshold t in terms of the mean and variance of the distribution.

From Eq. (10), we note that $\sigma^2 \leq \bar{\rho}(1 - u^2)$, this gives the upper bound of σ^2 . Based on $\sigma^2 \leq \bar{\rho}(1 - u^2)$, the Eqs. (10) and (11) can be expressed as follows:

$$P(Z \geq t) \leq \frac{1}{1 + \frac{(t-u)^2}{\bar{\rho}(1-u^2)}}, (t > u) \quad (12)$$

$$P(Z \geq t) \geq \frac{1}{1 + \frac{\bar{\rho}(1-u^2)}{(t-u)^2}}, (t < u) \quad (13)$$

For $t \in [u_0, 1]$ and $t = u_0 + k$, the FAR in this bound can be expressed as follows:

$$\text{FAR} = P(Z_0 \geq t) \leq \frac{1}{1 + \frac{(t - u_0)^2}{\bar{\rho}_0(1 - u_0^2)}} \quad (14)$$

Based on Eq. (4), (14) can be expressed as follows:

$$\text{FAR} = P(Z_0 \geq t) \leq \frac{1}{1 + \frac{(t + s_0)^2}{\bar{\rho}_0(1 - s_0^2)}} \quad (15)$$

In score distribution, u_0 and u_1 divide $[-1, 1]$ into three subintervals $[-1, u_0]$, $[u_0, u_1]$, and $[u_1, 1]$. Based on Eqs. (8), (12), (13), and (4), we give the bound of FAR and DET on three subintervals. The bounds are presented in Table 2.

Careful inspection of the bounds presented in Table 2 reveals the desired characteristics of the class-specific strength and correlation of the base classifiers that will yield bounds most favorable to ensemble performance. For example, in Region I, when strength is held fixed, it is clear that decreasing the correlation for the negative class samples $\bar{\rho}_0$ decreases the upper bound for the false alarm rate, potentially resulting in improved performance. Similarly, increasing the correlation for the positive class samples $\bar{\rho}_1$ will increase the upper bound of the detection rate. Though improved performance is not guaranteed, these bounds suggest guidelines for tuning the ensemble to produce more favorable conditions for minimizing class-specific errors.

It should be noted here that the region-specific guidelines derived from these bounds are highly consistent with the intuition gleaned from the score distribution diagram in Fig. 2. As we observed in Sect. 3.1, for a fixed strength, an increase in the class-specific correlation can lead to an increase in the variance of the corresponding score distribution.

As shown in Table 3, the error bounds for Regions I and III yield opposing guidelines with respect to class-specific mean correlation. Specifically, if the class-specific correlations could be effectively controlled for fixed means, performance within these regions of the true ROC curve could be explicitly traded off based upon relative error costs.

Figure 2 illustrates that when the decision threshold is very high in Region I, an increase in the spread (i.e., variance) of the class 1 score distribution, for a fixed mean, may increase the number of scores lying to the right of threshold, thus increasing the detection rate. This is a form of stochastic resonance, in which

Table 2 Bound of ROC

Region III $t \in [-1, -s_0]$	Region II $t \in [-s_0, s_1]$	Region I $t \in [s_1, 1]$
$\text{FAR} \geq \frac{1}{1 + \frac{\bar{\rho}_0(1-s_0^2)}{(t+s_0)^2}}$	$\text{FAR} \leq \frac{1}{1 + \frac{(t+s_0)^2}{\bar{\rho}_0(1-s_0^2)}}$	$\text{FAR} \leq \frac{1}{1 + \frac{(t+s_0)^2}{\bar{\rho}_0(1-s_0^2)}}$
$\text{DET} \geq \frac{1}{1 + \frac{\bar{\rho}_1(1-s_1^2)}{(t-s_1)^2}}$	$\text{DET} \geq \frac{1}{1 + \frac{\bar{\rho}_1(1-s_1^2)}{(t-s_1)^2}}$	$\text{DET} \leq \frac{1}{1 + \frac{(t-s_1)^2}{\bar{\rho}_1(1-s_1^2)}}$

Table 3 Tuning class-specific correlation

Region	Guideline
I	$\downarrow \bar{\rho}_0$ and $\uparrow \bar{\rho}_1$
II	$\downarrow \bar{\rho}_0$ and $\downarrow \bar{\rho}_1$
III	$\uparrow \bar{\rho}_0$ and $\downarrow \bar{\rho}_1$

adding variability to the system improves performance. Intuitive arguments similar to that above can be made regarding the bounds in the remaining regions. Note that in all cases when correlation is held fixed, higher strength for both classes produces a greater separation of the score means and may yield improved performance. For a fixed strength, the guidelines for tuning class-specific correlation inferred from the bounds for each region are summarized in Table 3.

3.4 The Procession of IRFA

Because Regions I and III correspond to low false alarm and missed detection rates, respectively, they are of great interest for the many real-world applications that involve extreme differences in error cost. Like Breiman's bound, the error bounds derived for these regions are relatively loose; hence, they serve most effectively as an intuitive guide to performance optimization. For medical image classification, it is great to improve the DET and decrease the FAR as far as possible.

We give the procession of IRFA as follows:

- (1) Import data: put the data that obtained from Section 2 into database.
- (2) Extraction of training set and test set: give the ratio of training samples and test samples as $N:M$, select $N/(N+M)$ from the all set as training samples, the other as test samples.
- (3) Set the parameter: set the number of training trees and split dimension.
- (4) Construct the trees: it is the core of RF.

We use bagging method to select the samples and use CART as the base classifier.

- (5) The evaluation of classification performance.

We give the test of RF based on the RF and give the evaluation from different class. Based on the guidelines of Table 3, we retraining the RF.

4 Experiment and Analysis

In the paper, the medical image we used is real data from tumor hospital. The corresponding diagnosis records we generalized to negative (normal) and positive (abnormal). We collect 1,000 images and get 15 features, we call the ability of

Table 4 Performance of different number trees

Number of trees	Accuracy (%)	DET (%)	FAR (%)
51	84.9	85.1	15.5
75	87	87.7	13.7
101	88.6	88.9	11.7
151	88.6	89.2	12
201	88.6	89	11.8
251	88.6	89	11.8

Table 5 Performance of different split dimension

m	Accuracy (%)	DET (%)	FAR (%)
2	84.9	85.1	15.5
3	87	87.7	13.7
4	88.6	88.9	11.7
5	88.6	89.2	12
6	88.6	89	11.8

identify positive sample as DET and FAR as the ability of misclassified negative sample. We use 10-fold cross-validation begin the experimental test. Firstly, we give the test of the selection of the number of trees as shown in Table 4.

From Table 4, we select the number of trees as 151.

Then, we give the test of the selection of the number of split dimension (we called m) as shown in Table 5.

From Table 5, we select the number of split dimension as 5.

As we select 151 as the number of trees and 5 as the number of split dimension, we give the ROC under the different threshold for classification as shown in Fig. 3. We select the threshold as 1, 0.8, 0.5, 0, -0.5 , -1 .

From Fig. 3, we select the threshold as 0.5 in our experiment.

As we select 151 as the number of trees, 5 as the number of split dimension, and 0.5 as the threshold, we give the compare experiment as shown in Table 6.

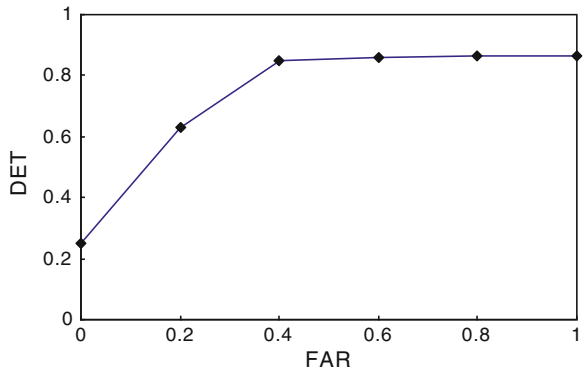
Fig. 3 Figure of ROC

Table 6 Compare experiment

Method	Accuracy (%)	DET (%)	FAR (%)
IRFA	88.8	89.3	11.7
RF	87.3	87.4	12.8

From Table 6, IRFA gives better performance than RF on medical image classification.

5 Conclusion

In this paper, we propose IRFA to solve the classification problem. It leverages key elements of the derivation of generalization error bound to derive bounds on DET and FAR on ROC and gives the performance optimization guidelines for tuning class-specific correlation inferred from the bounds for each region. Experiments show that IRFA gets fast convergence performance and efficient classification performance at accuracy, and DET and FAT are relatively low.

Acknowledgments The paper is partly supported by the National Natural Science Foundation of China under Grant No. 61272184, 61202090; Natural Science Foundation of Heilongjiang Province under Grant No. F200903, F201016, F201024, F201130; the Program for New Century Excellent Talents in Universities (NCET-11-0829); the Fundamental Research Funds for the Central Universities under grant No. HEUCFZ1010.

References

1. Zaiane OR, Han JW, Li ZN, Hou J, Liu G (1998) Mining multimedia data. Proceedings CASCON'98: meeting of minds, Toronto, Canada, pp 83–96, Nov 1998
2. Pan HW, Zhang N, Han QL, Yin GS (2010) Incorporating domain knowledge into multi-strategical image classification. In: The 2nd international workshop on database technology and application, pp 399–402
3. Sheng VS, Ling CX (2006) Thresholding for making classifiers cost-sensitive. American Association for artificial intelligence, pp 476–481
4. Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32
5. Pan HW, Li JZ, Zhang W (2007) Incorporating domain knowledge into medical image clustering. *Appl Math Comput* 185(2):844–856
6. Ester M, Kriegel HP, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases. *KDD'96*, pp 118–127
7. Rokach L (2009) Taxonomy for characterizing ensemble methods in classification tasks: a review and annotated bibliography. *Comput Stat Data Anal* 53:4046–4072
8. Garg A, Pavlovic V, Huang TS (2002) Bayesian networks as ensemble of classifiers. In: 16th international conference on pattern recognition (ICPR'02), 2, pp 779–784
9. Koltchinskii V, Panchenko D, Lozano F (2003) Bounding the generalization error of convex combinations of classifiers: balancing the dimensionality and the margins. *Ann Appl Probab* 13(1):213–252

IdeaGraph: Turning Data into Human Insights for Collective Intelligence

Hao Wang, Yukio Ohsawa, Pin Lv, Xiaohui Hu and Fanjiang Xu

Abstract Data mining has been widely applied for business data analysis, but it only reveals a common pattern based on large amounts of data. In current years, chance discovery as an extension of data mining is built to detect rare but important chances for human decision making. KeyGraph is an algorithm as well as a tool for discovering rare and important events that are regarded as candidates of chances in chance discovery. However, KeyGraph is originally invented as a keyword extraction algorithm, so scenario graph generated from KeyGraph is machine oriented, which causes a bottleneck of human cognition. Traditional data mining methods also have the similar problem. In this paper, we propose a user-oriented algorithm called IdeaGraph which can generate a rich scenario graph for humans' comprehension, interpretation, and innovation. IdeaGraph not only works on discovering more rare and significant business chances, but also focuses on uncovering latent relationship among them. An experiment indicates the advantages and effects of IdeaGraph by comparing with KeyGraph. IdeaGraph has been integrated in a creativity support system named iChance for collective intelligence.

Keywords IdeaGraph · KeyGraph · Chance discovery · Human insight

H. Wang (✉) · P. Lv · X. Hu · F. Xu
State Key Laboratory of Integrated Information System Technology, Institute of Software,
Chinese Academy of Sciences, 100190 Beijing, China
e-mail: wanghao0423@gmail.com

H. Wang · Y. Ohsawa
Department of Systems Innovation, The University of Tokyo, Tokyo 113-8656, Japan

1 Introduction

Chance discovery is a human–computer interaction process that detects rare but important chances for decision making. A chance is defined as an infrequent but significant event or situation that strongly impacts on human decision making [1, 2]. In the computer process, collected data are analyzed and visualized into a scenario graph by a tool named KeyGraph [3]. Humans become aware of chances shown in the scenario graph and explain their significance, especially if the chance is rare and its significance is unnoticed [2, 4].

To improve human insight into scenario graph, a group-based discussion is designed to support the interaction process between human and scenario graph [4, 5]. In recent years, Innovators’ market game (IMG), a tool of chance discovery, has been developed for innovative thoughts and communication [6]. Two customer-centric innovation approaches, 4 W-IMG and Market innovation storming (MIS) are further developed for high-quality idea generation. iChance, a Web-based creativity support system, has been built and integrated with 4 W-IMG and MIS approaches for collaborative innovation, such as idea generation, evaluation, and knowledge creation [7–9].

KeyGraph is originally a graph-based keyword extraction algorithm [10] and is gradually used to analyze a variety of data and find rare or novel events in chance discovery. It has achieved successful outcomes in detecting earthquake risks [11], discovering emerging topics from the WWW [12], analyzing financial trends [13, 14], seeking new hit products [15], and triggering creative ideas [7–9, 16–19]. However, scenario graph from KeyGraph is machine oriented, which causes a bottleneck of human cognition.

KeyGraph has been applied in many areas to help users solve problems and learn how to understand scenario graph through group discussion, group meeting, and innovation game. However, no work has yet been attempted to develop a new algorithm that generates a scenario graph to obtain more effective human insights. Therefore, a problem needs to be addressed that is “Is there a method that may discover latent chance from data as well as turn data into human insights?”

This paper concentrates on a new algorithm to improve human insights by modeling data as a network. A review of KeyGraph algorithm is made in Sect. 2, and its limitation is illustrated in Sect. 3. To overcome the problems of KeyGraph, a new algorithm named IdeaGraph is presented in Sect. 4. An experiment is performed to validate the effects of IdeaGraph in Sect. 5, and the conclusion is summarized in Sect. 6.

2 A Review of KeyGraph Algorithm

KeyGraph is a keyword extraction algorithm based on co-occurrence graph from a single document. In KeyGraph, a document is visualized as a two-dimensional undirected graph where each node corresponds to a term (word or phrase) and each

edge represents co-occurrence terms. Based on the segmentation of graph into clusters, keywords are extracted by selecting the terms which strongly co-occurs with multiple clusters.

A document consists of a list of sentences, and each sentence has a series of words or phrases. So a document is firstly preprocessed into a bag of terms D as the input of KeyGraph by stemming and removing stop words.

$D =$ (Sentence 1) term 1, term 2, term 3, term 4.
 (Sentence 2) term 2, term 7, term 5.
 (Sentence 3) term 3, term 6, term 9, term 10, term 5.

- Step 1 Extracting high-frequency terms. Terms are sorted by occurrence frequency in the data D . High-frequency terms are picked up and denoted as black nodes in the graph G
- Step 2 Extracting strong co-occurrence term-pairs in high-frequency items. The co-occurrence of term-pairs is calculated by the Jaccard coefficient in Eq. (1)

$$J(I_i, I_j) = \frac{P(I_i \cap I_j)}{P(I_i \cup I_j)} \tag{1}$$

where I_i and I_j are high-frequency terms, $P(I_i \cap I_j)$ is the probability of term I_i , and I_j co-occurring in the same sentence, $P(I_i \cup I_j)$ is the probability of the term I_i or I_j occurring in the sentence.

The strong co-occurrence item pairs are taken to link with black solid line in the graph G , thus clusters are formed.

- Step 3 Extracting high-key terms. High-key term is the term that strongly co-occurs with clusters in the graph G . The key value of each term in the data is calculated by Eq. (2).

$$\text{key}(I) = 1 - \prod_{C \subset G} [1 - J(I, C)] \tag{2}$$

The high-key terms are added with red nodes if they are not in the graph G . Each high-key item is linked to two or more clusters by red dot lines. These high-key items like a hub bridging different clusters are regarded as candidates of chance that could be important for decision making. Figure 1 shows an example of a scenario graph from KeyGraph.

3 The Limitation of KeyGraph

By analyzing the KeyGraph algorithm, we found some limitations as below:

1. *It disregards parts of key items as candidates of chance.* The $J(I,C)$ calculation of a term co-occurring with other clusters in Eq. (2) does not take the basket data's capacity (the number of terms in each sentence) into account. If the number of terms in a cluster is equal to or greater than the basket data's capacity, $J(I,C)$ of any term outside the cluster is 0. So some key terms that probably have strong co-occurrence with subset of a cluster will be filtered out. For instance, suppose in Fig. 1 that term 3 has strong co-occurrence with two subsets: {term 7, term 2} in Cluster 1 and {term 1, term 5} in Cluster 2, and the capacity of all basket data including term 3 is less than 3, so $J(\text{term 3, Cluster 1})$ and $J(\text{term 3, Cluster 2})$ is equal to 0 and term 3 cannot be shown in Fig. 1 as a bridge connecting Cluster 1 and Cluster 2. In fact, term 3 is still a potential key candidate of a chance
2. *It cannot find strong relationship between low-frequent terms.* The low frequent terms probably have high co-occurrence in the data. As Fig. 2 shows, low-frequent terms 3 and 8 have high Jaccard efficient
3. *It fails to capture low-frequency but important terms in the clusters.* As shown in Fig. 2, a low-frequent term (4 or 11) as a hub that links other terms in the same cluster should be treated as a key term. In general, a company's executive

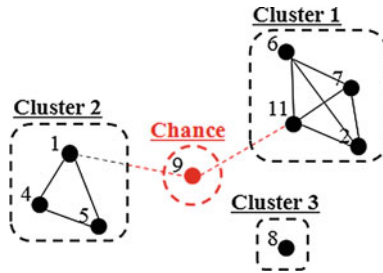


Fig. 1 An example of scenario graph as the output of KeyGraph

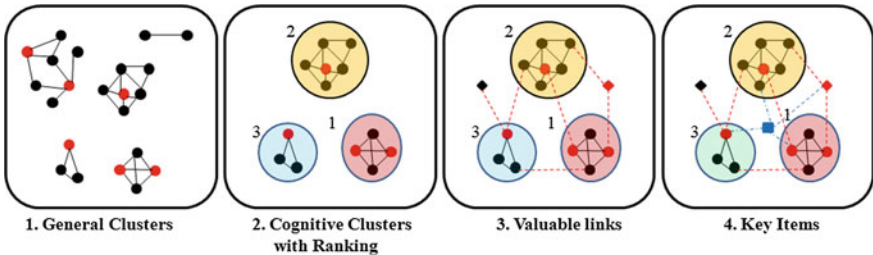


Fig. 2 Scenario graph formation process using IdeaGraph algorithm

does not appear very often, but he or she has strong connections with each head of the departments

- 4. *It ignores direct links between the clusters.* A chance is not necessarily an event point (a term). A direct link between different clusters should also be considered as a chance, see Fig. 2. In other words, a situation (Cluster 1) transferring to another situation (Cluster 2) may take direct links between clusters rather than an event point outside the clusters. The former probably would be more effective than the latter
- 5. *It cannot discover the relationship between chance items.* There might be latent structure between chance items. In other words, the connections between red nodes cannot be achieved in KeyGraph.

KeyGraph algorithm is originally designed for extracting keywords in a document. The document is represented as scenario graph, and the terms co-occurring with multiple clusters are selected as keywords. But sometimes scenario graph from KeyGraph is hard for users to understand and interpret because of its complexity and inadequate information. Thus, based on document or business data, IdeaGraph is developed to generate a scenario graph which can obtain more effective and deeper human insights.

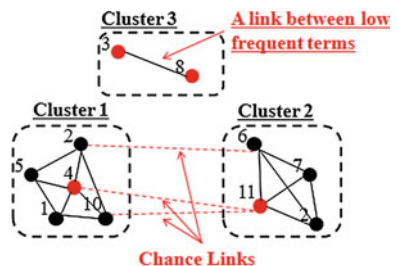
4 IdeaGraph: A New Algorithm to Obtain Improved Insights

Suppose that data have been preprocessed into D'

D' = item 1, item 2, item 3, item 4
 item 2, item 7, item 5
 item 3, item 6, item 9, item 10, item 5

Figure 3 shows a scenario graph generation process by IdeaGraph, the algorithm of which is presented as below:

Fig. 3 Valuable links and key items that KeyGraph fails to visualize



Step 1 *Generating general clusters.* The relationship between two items is measured by their conditional probability. That is, the relationship of any two items, I_i and I_j , is calculated by Eq. (3)

$$R(I_i, I_j) = P(I_i|I_j) + P(I_j|I_i) \quad (3)$$

Then the pairs whose $R(I_i, I_j)$ are greater than preset threshold r are linked by line in the graph G . Finally, general clusters emerge and are denoted by C_i .

Step 2 *Obtaining cognitive clusters.* Cognitive Cluster is defined that a cluster embraces rich information but should be small enough for human cognition and interpretation.

To obtain cognitive cluster, two indicators, *information and information density*, are employed to quantify general clusters generated in Step 1. The definition of *information* is the sum of $R(I_i, I_j)$ of all the edges in a general cluster. The *information density* is defined that the *information* of a cluster is divided by the number of items in the cluster. That is, the *information density* of a cluster is the information of each item in this cluster. Thus, the equations of *information* and *information density* are

$$\text{Info}(C) = \sum_{I_i, I_j \in C} R(I_i, I_j) \quad (4)$$

$$\text{InfoDen}(C) = \frac{\text{Info}(C)}{N_e} \quad (5)$$

where I_i or I_j is an item of a cluster C and N_e indicates the number of items in the cluster C .

Therefore, each general cluster is measured by the harmonic average of these two indicators. Equation (6) is derived by merging Eqs. (4) and (5). Finally, the value of each general cluster is measured by the harmonic average of these two indicators, see Eq. (6).

$$\text{ClusterVal}(C) = \frac{2\text{Info}(C)}{N_e + 1} \quad (6)$$

Equation (6) indicates when two general clusters have the same *information*, it favors the cluster that has fewer items.

Therefore, all general clusters are ranked by their $\text{ClusterVal}(C)$ in a descending order, and parts of them are chosen as cognitive clusters denoted by CC through picking up the top N_c clusters.

Step 3 *Capturing valuable links.* Calculate the relationship between each item and each cognitive cluster by Eq. (7)

$$PR(I_i, CC) = \sum_{I_i \notin CC, c_k \in CC} R(I_i, c_k) \quad (7)$$

where c_k is an item of a cognitive cluster CC and I_i is an item outside the cluster CC .

Then, item-cluster pairs are sorted, and top M_1 pairs are selected to be linked by red dot line. New items are added if they are not in the graph G .

Step 4 *Extracting key items*. A key item is the item that has strong relationship with all other cognitive clusters and newly added items in Step 3. It is calculated by Eq. (8).

$$\text{Key}(I) = \sum_{i=0}^{N_c} PR(I, CC_i) + \sum_{I_k \notin CC, I_k \in G} R(I, I_k) \quad (8)$$

All items are sorted by their $\text{Key}(I)$, and top M_2 items are taken as key items that are shown if they do not exist in the graph G .

5 Experimental Evaluation

In this section, we report the result of an experiment to demonstrate the advantages of IdeaGraph by comparing them with the result achieved using KeyGraph.

5.1 Experimental Design

The task of our experiment is to assist an automobile manufacturer to find out Chinese customers' preference on a car system. Figure 4 shows that the experiment is performed using the data from 109 questionnaires of automotive customers. Firstly, the data from questionnaires are preprocessed into 714 basket data sets where each item represents a specific function of the car system, such as navigation, telephone, video/audio, entertainment information, etc. Secondly, two scenario graphs are separately generated by KeyGraph and IdeaGraph. Finally, we compare two scenario graphs and evaluate the effect of IdeaGraph.

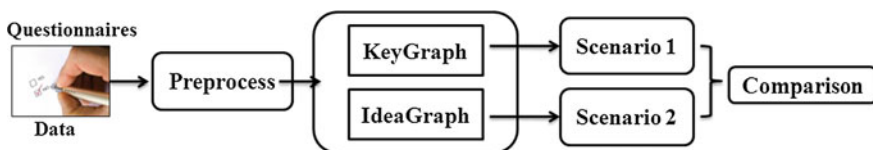


Fig. 4 Experimental procedure for comparison

5.2 Results' Evaluation

Figure 5 shows *Scenario 1* from *KeyGraph*, and Fig. 6 indicates *Scenario 2* generated by *IdeaGraph*. The parameters of *KeyGraph* and *IdeaGraph* are set in a comparable standard since these two algorithms have different parameter settings. Table 1 indicates *Scenario 2* presents more information than *Scenario 1*. Moreover, additional information in *Scenario 2* can be well understood and interpreted.

According to the limitation of *KeyGraph* presented in Sect. 3, we present advantages of *IdeaGraph*.

1. *IdeaGraph* can uncover more key items. More key items (red nodes) are shown in Fig. 7a, such as *3D display technology*, *multi-touch screen*, and *voice recognition technology*. These are potentially important demands of car holders with the development of information technology
2. *IdeaGraph* may discover strong relationship between low frequent items. Figure 7b shows two diamond nodes in Cluster 4, and *screen display* and *conflict of driver and front-row passenger* are low frequent but have strong relationship. There exists a strong demand that a driver and a passenger in the front seat both can use screen. The red node of *Dual-View Screen Technology* is a good solution to that demand

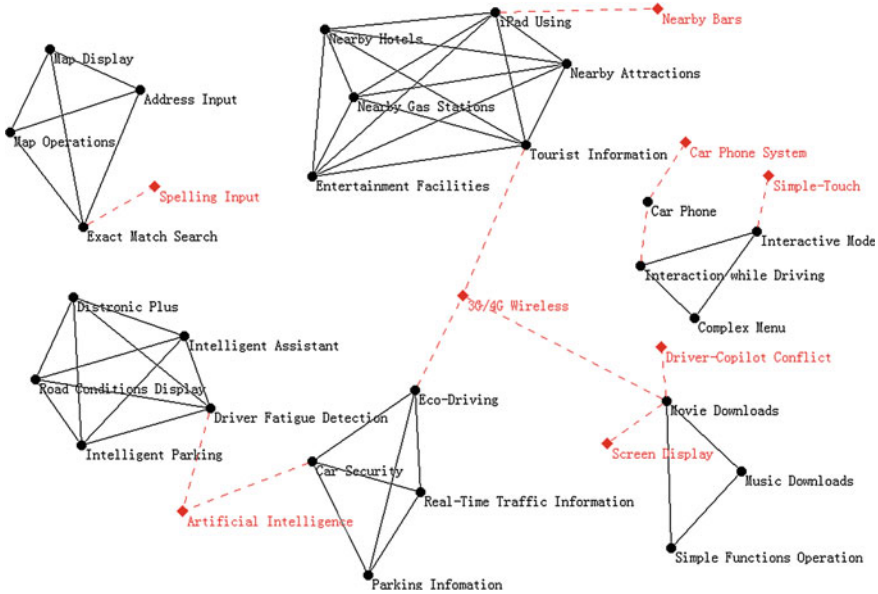


Fig. 5 Scenario 1 created by KeyGraph

Table 1 The comparison of KeyGraph and IdeaGraph

Algorithm	Key items	Valuable links
KeyGraph	8	12
IdeaGraph	15	41

- IdeaGraph can capture low-frequency but important items inside the clusters, such as nearby bars in Cluster 1, spelling input in Cluster 3, and simple touch in Cluster 6, see Fig. 7c. Although these nodes do not appear very often, they have strong relationship with other nodes inside the cluster*
- IdeaGraph can find direct links among clusters, such as road conditions display in Cluster 2 and map display in Cluster 3, shown in Fig. 7d. The direct link indicates users need more road information displayed in the map*
- IdeaGraph is able to find potential structure among chance items, such as the link between 3D display technology and multi-touch screen, shown in Fig. 7e. More and more users hope 3D and Touch Screen are equipped in their car, which will bring them a lot of fun and better user experience.*

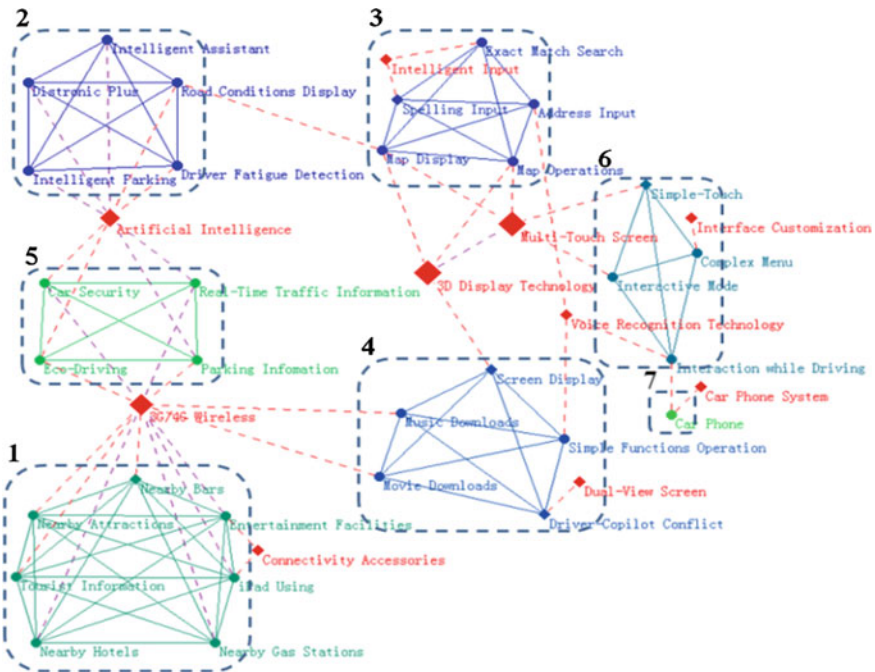


Fig. 6 Scenario 2 generated by IdeaGraph

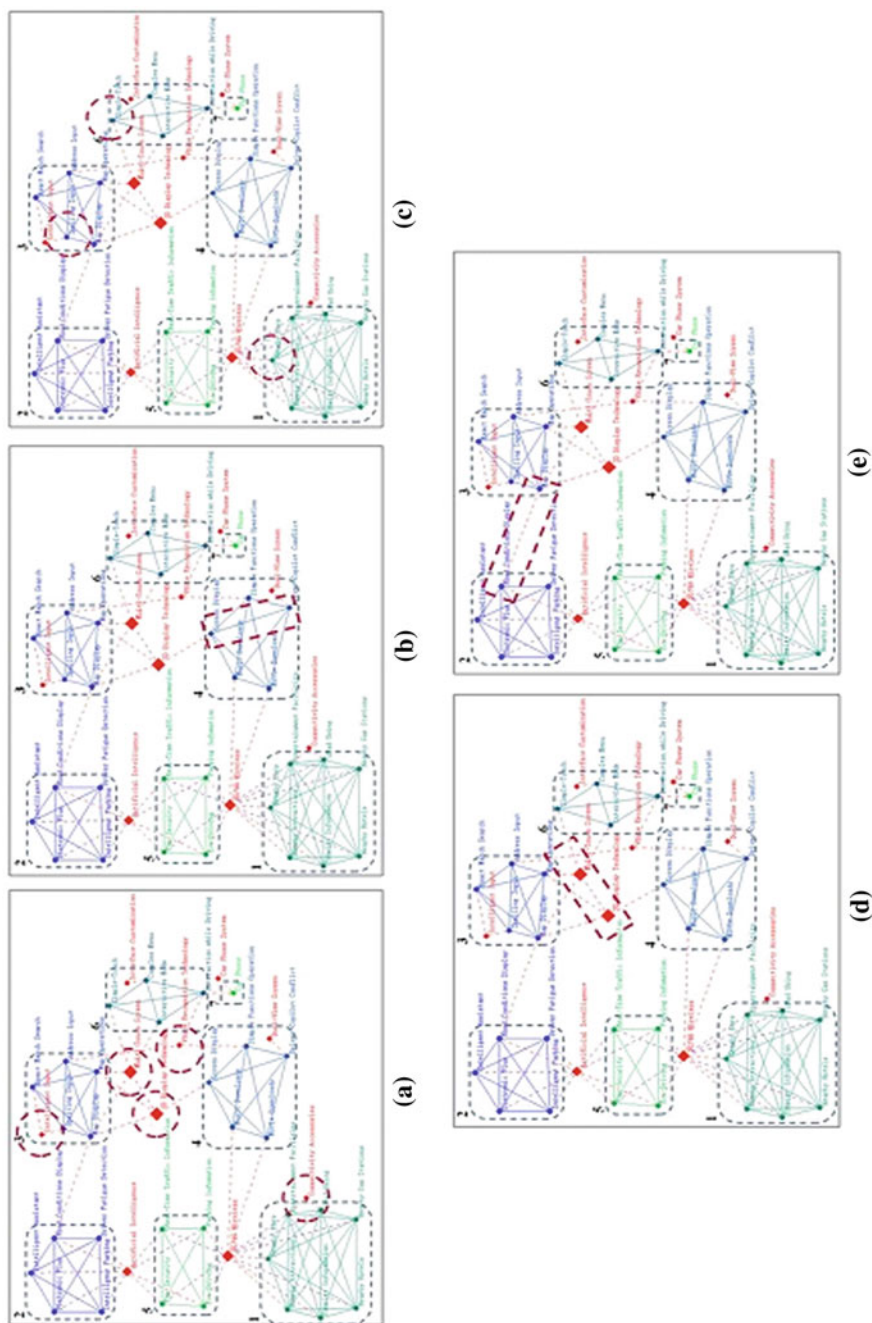


Fig. 7 The advantages of IdeaGraph

6 Conclusion

In this paper, we propose a new algorithm IdeaGraph which can discover latent information and generate rich scenario graph for human insights. In IdeaGraph, the definition of a chance is extended that a rare and important relationship among events is still regarded as a chance. An experiment indicates the advantages of IdeaGraph by comparing with KeyGraph that IdeaGraph can uncover more key events, find out strong relationship between rare events, capture rare but significant events inside the clusters, and discover latent structure among key events. Moreover, IdeaGraph is easier for humans to comprehend, interpret, and innovate. We have applied it in a Web-based creativity support system called *iChance* for collective intelligence [7–9].

Acknowledgments The author was supported through the Global COE Program “Global Center of Excellence for Mechanical Systems Innovation,” by the Ministry of Education, Culture, Sport, Science, and Technology of Japan and was also funded by Chinese Scholarship Council (CSC).

References

1. Ohsawa Y (2002) Chance discoveries for making decisions in complex real world. *New Gener Comput* 20(2):143–163
2. Ohsawa Y (2003) Mcburney P (eds) *Chance discovery*. Springer, Berlin
3. Ohsawa Y (2005) Data crystallization: chance discovery extended for dealing with unobservable events. *New Math Nat Sci* 1(3):373–392
4. Ohsawa Y, Fukuda H (2002) Chance discovery by stimulated groups of people. Application to understanding consumption of rare food. *J Contingencies Crisis Manage* 129–137
5. Fukuda H, Ohsawa Y (2001) Discovery of rare essential food by community navigation with KeyGraph—an introduction to data-based community marketing. In: *Proceedings of the fifth international conference on knowledge-based intelligent information and engineering systems (KES2001)* IOS Press, Osaka, pp 946–950
6. Ohsawa Y, Okamoto K, Takahashi Y, Nishihara Y (2010) Innovators marketplace as game on the table versus board on the web. In: *Proceeding of the IEEE international conference on data mining workshops*, Sydney, Australia, pp 816–821
7. Wang H, Ohsawa Y (2011) *iChance*: a web-based innovation support system for business intelligence. *Int J Organ Collective Intell* 2(4):48–61
8. Wang H, Ohsawa Y (2011) *iChance*: towards new-generation collaborative creativity support system for advanced market innovation. In: *Proceeding of the 6th international conference on knowledge, information and creativity support systems (KICSS)*, Beijing, China, pp 138–143
9. Wang H, Ohsawa Y, Nishihara Y (2011) Web-based innovation supporting system for creative ideas emerging. In: *Proceedings of the 6th international workshop on chance discovery in the 22nd international joint conference on artificial intelligence (IJCAI-11)*, Barcelona, Spain, pp 15–20
10. Ohsawa Y, Benson NE, Masahiko Y (1998) KeyGraph: automatic indexing by cooccurrence graph based on building construction metaphor. In: *Proceeding of advanced digital library conference (IEEE ADL'98)*, pp 12–18
11. Ohsawa Y (2002) KeyGraph as risk explorer in earthquake-sequence. *J Contingencies Crisis Manage* 10(3):119–128

12. Matsumura N, Matsuo Y, Ohsawa Y, Ishizuka M (2002) Discovering emerging topics from WWW. *J Contingencies Crisis Manage* 10(2):73–81
13. Hong CF, Chiu TF, Chiu YT, Lin MH (2007) Using conceptual scenario diagrams and integrated scenario map to detect the financial trend. In: *Proceedings of the 20th international conference on industrial, engineering and other applications of applied intelligent systems*, pp 886–895
14. Chiu TF, Hong CF, Chiu YT (2008) Visualization of financial trends using chance discovery methods. In: Nguyen NT et al (eds) *Lecture notes in artificial intelligence. Proceedings of IEA/AIE 2008*, vol 5027. pp 7080–7717
15. Usui M, Ohsawa Y (2003) Chance discovery in textile market by group meeting with touchable KeyGraph. In: *Proceedings of social intelligence design*
16. Wang H, Ohsawa Y, Nishihara Y (2011) A system method to elicit innovative knowledge based on chance discovery for innovative product design. *Int J Knowl Syst Sci* 2(3):1–13
17. Wang H, Ohsawa Y (2011) Innovation support system for creative product design based on chance discovery. *Expert Syst Appl* 39:4890–4897
18. Ohsawa Y, Okamoto K, Takahashi Y, Nishihara Y (2010) Innovators marketplace as game on the table versus board on the web. In: *Proceeding of the IEEE international conference on data mining workshops, Sydney, Australia*, pp 816–821
19. Ohsawa Y (2009) Innovation game as a tool of chance discovery. In: *Proceedings of the 12th international conference on rough sets, fuzzy sets, data mining and granular computing*, pp 59–66

A General Hierarchy-Set-Based Indoor Location Modeling Approach

Jingyu Sun, Huifang Li and Hong Huang

Abstract Most indoor business processes are based on the locations of physical objects, so building a well-formed model to represent the spatial knowledge about business process is necessary under ubiquitous computing environments. Current location modeling requires a large amount of manual effort and cannot balance well between multipurpose query and less cost. In this paper, having analyzed the requirements of typical location-based services and the suitability of existing location modeling approaches, we propose a combined indoor location model which extends an improved “directly under” relation-based hierarchy tree model by introducing a core set model, based on the “level reachable” relation. Through a series of modeling theoretical principles and a real example, we show that our model is simple but indoor general enough to capture both spatial connectivity and containment relationships and support more location-based applications. Furthermore, a single-location hierarchical tree model, which integrates room coordinates, the building exits, and corridor intersections together, can navigate more source and objective positions in reality and provide more than one candidate path corresponding to the changing environment. At the same time, our combined model can be more flexible to be integrated into a context-aware system model.

Keywords Indoor location model · Location-based services · Hierarchy · Core set · Path planning

J. Sun (✉) · H. Li · H. Huang
Automation Department, Beijing Institute of Technology University, Beijing, China
e-mail: jingyu1968110@163.com

H. Li
e-mail: huifang@bit.edu.cn

H. Huang
e-mail: honghuang@bit.edu.cn

1 Introduction

As a new generation of computing pattern after the mainframe and desktop computing, pervasive computing [1], which was proposed by Mark Weiser in 1991, aims to weave computation into daily life and enable users to automatically get dynamic information and services anytime and anywhere. It is upon this background that context-aware system [2] has been developed and spread gradually. Context is any information that can be used to characterize the situation of an entity including the users and applications themselves [3]. Among various sources of information, location is commonly considered to be one of the primary contexts used to indexing the secondary ones, especially in the LBS (location-based service) domain, which plays a more and more important role in office automation and business management.

There are many kinds of LBS in need, such as location identifying, browsing, navigation, and querying. It is required to provide the end users with a formalized and semantic rich locations knowledge representation and applications. Furthermore, it should be understandable and convenient to communicate among location sensors, database servers, and end users. This paper presents a location modeling approach to realize these objectives indoor.

Firstly, this paper analyzes the general requirements on location management and then enumerates the existing location modeling approaches according to their suitability to various location-based query services. Secondly, through improving and extending the semantic location model in [4], a more general indoor location model approach is proposed in this paper. Our model comprehensively considers various position nodes, including room coordinates, building exits, and corridor intersections, and packs them into a single hierarchical location tree, so that it not only can be more accurate and reliable than the indirectly deduced area model based on an only exits included model, but also can navigate to more source and objective positions, instead of merely room nodes navigation. This accords more with the practical situation and is exactly necessary in the real world.

To provide range querying, other than just such reachable topological semantic-based applications as nearest querying and navigation, the set-based modeling approach will be combined to the hierarchy-based model. By making full use of the advantage of the hierarchical structure in efficient classifying and searching, we just need to create and store some core sets based on a “level reachable” relation, and then, other subsets will be constructed automatically when needed, while in the existing set-based location models, a large amount of symbolic coordinate subsets for some certain ranges within the universal area to be modeled have to be predefined manually and cannot be used for navigation. In a word, the new location model approach has a combination of wide uses and low costs. Finally, a case study shows that the proposed hierarchy-set combined modeling method is feasible and flexible for the indoor location-based mobile and ubiquitous application development.

The rest of the paper is organized as follows: [Sect. 2](#) is an overview of the applications and requirements of location modeling. In [Sect. 3](#), we improve and extend a hierarchical semantic location modeling approach and then propose a new path planning algorithm based on the new hierarchy-set combined model. [Section 4](#) is a modeling example analysis, and [Sect. 5](#) presents a conclusion and future research.

2 Overview of Location Modeling

2.1 Location-Based Services

Considering the users' activities and most applications which involve location information, four kinds of location-based queries have been summarized [5] in order to further derive the functional requirements of location models and model properties with respect to different organizations.

2.2 Position Query

Determining the position of mobile and static objects, like users, buildings, bus stops, etc., is a basic function module of location-based service and context-aware application systems. And the query tasks to be described below cannot be carried out without knowing the objects' position information. Thereby, all location models should contain this information and represent it with various forms of coordinate.

Different situational applications have to choose their best suitable coordinate systems. For example, in a smart factory, there always is a production plan that requires a common interpretation of the coordinates in a specific global coordinate system so as to monitor the positions of resources and tools. While within moving objects, such as trains, local reference systems with respect to their compartment in the train other than their absolute position to the ground can better help to address objects or travelers.

The existing coordinate systems fall into two basic classes: geometric and symbolic coordinates [6].

Geometric coordinates They define position in the form of tuples relative to some given reference coordinate systems. Among them, the World Geodetic System 1984 (WGS84) is a global reference system and thus can be used to mark places on the planet by a triple group with longitude, latitude, and altitude, whereas the Cartesian coordinates of the active bat system [7], which provide three-dimensional coordinates, that is, x , y , z value for high-resolution indoor positioning, are typically only valid locally. Geometric location models can be used to

more accurately calculate the distance (in a straight line) between two geometrically defined positions. At the same time, topological relations, like spatial containment and simple spatial reasoning, can be derived from the geometry characteristics of objects. Remarkably, the “connected to” relation modeling, for example, doors connecting rooms, cannot be derived from such static map database with location geometries like GIS, and many practical LBS would not be completely realized only with the geometric models.

Symbolic coordinates They define geographical areas in the form of abstract symbols, for example, room and street names, which do not need to be constant over time. The active badge [8] system provides unique symbolic identifiers for locations via IR sensors registering the users’ active badges. Symbolic location models provide additional information about symbolic coordinates whose types may be set-based, hierarchical, graph-based, or combined, so that they can describe not only the topological relation of “contained in,” but also the “connected to” relation between nodes.

In contrast to geometric coordinates, the distance between two symbolic coordinates is not implicitly defined. Also, topological relations like spatial containment cannot be determined without information about the relationships between symbolic coordinates.

2.3 Nearest Service Query

Nearest service query is the search for some objects closest to a certain location, like the nearest restaurant or the next printer. Object positions and the reachable distance between two coordinates are required for this type of query. It is notable that there are other notions of distance on the coordinates which are often more relevant than the direct physical distance. For instance, for a pedestrian, it is always impossible to cross a highway. Therefore, a restaurant across the highway with a linear distance of 100 m might be farther to reach than another with 200 m linear distance not located across this highway. In these cases, additional model information, like the road network or the length of path from location A to location B, has to be taken into account.

2.4 Navigation Query

As people put forward more requirements for the traffic convenience and quickness of the traffic higher, navigation systems have become standard equipment in most modern cars. A location model, which is used to look for paths between locations with the information of transportation network (roads, train, or bus routes, etc.) and interconnected topology relations, would help to realize such systems.

2.5 Range Query

A range query returns all objects within a certain geographic area. The typical applications include a correctly processed evacuation plan that ensures the room is empty before the fire doors are closed and real-time communication implemented with simpler algorithms, etc. To implement a range query, object positions and the topological relation “contains” have to be modeled.

2.6 Location Semantics

From the use cases and corresponding requirements shown above, we can see that in general a location model needs not only to describe position coordinates, but also to provide at least the following two abstract location-based semantics: topology relation and distance.

2.7 Topology Relation Semantics

For range querying In the set-based symbolic models or geometric models, the “disjoint to,” “contained in,” and “overlapped to” relations between location nodes should be described to support range queries.

For path planning By building graph-based location models, another topology semantic is presented, that is, “connected to” or “reachable” relation. And these relations are necessary to both the navigation and nearest service query.

2.8 Distance Semantics

As previously mentioned, the definitions of “distance” differ between various practical applications. We may see the Euclidean distance in a location-based commercial advertising transmission application or the road-network distance in a finding nearest restaurant LBS application or even in some cases, the distance can be measured by the energy consumption over it. Actually, distance can be described qualitatively or quantitatively.

Except the above-mentioned position, topology relation, and distance semantics, a mobile object sometimes also needs the relevant orientation semantic for navigation from one location to another. For instance, there are two ways to pass through a cyclic gallery, the clockwise-based and counterclockwise-based which lead to different path distances. Getting to somewhere by bus, we must firstly figure out the principle direction and then determine to wait for bus at the right roadside in case of going far away from the destination.

2.9 Existing Location Modeling Approach

All the above-mentioned model elements and its technical parameter requirements, such as accuracy, updating rate, granularity and scope, should be comprehensively considered in modeling process, in order to make the model complexity as lower as possible on the premise of satisfying the model functional requirements. The properties of existing location models in supporting different kinds of queries are shown in Table 1. Each existing modeling approach has its corresponding implementations enumerated as follows:

- The active badge system uses set-based location model;
- An aware home model [9] for smart space is graph-based;
- EasyLiving system describes the spatial inclusion between locations hierarchically;
- The active map combines the benefits of graph-based and hierarchical location;
- A hybrid location model based on symbolic geometric coordinates for fine-grained regionally dissemination is adopted in Nexus system [10].

To avoid a high modeling complexity, not all of these semantics and factors have to be simultaneously fulfilled in most real applications. Therefore, being aware of the situation requirements and choosing the most appropriate location model structure are very important.

3 Combined Hierarchy-Set Model for Indoor Locations

Considering the model generality for various indoor layouts and applications, we adopt a combined model with taking the hierarchy and set into consideration, through which all the position elements are heterogeneously queried as they are in the real world, such as buildings, rooms, and corridors, and then, the tree elements are divided into some core sets later.

Table 1 Properties of existing location models

Location model		Supported queries			Model effort
		Position	Range	Nearest and navigation	
Symbolic coordinate based	Set-based	Good	Good	Basic	High
	Graph-based	Good	Basic	Good	Middle
	Hierarchy	Good	Good	Basic	Middle
	Combined	Good	Good	Good	High
Symbolic and geometric coordinate-based hybrid model		Good	Good	Good	Very high

3.1 Building Hierarchy Tree Model of Location

Location-related notions

Before introducing our modeling approach, the following location-related notions should be defined:

Definition 1 The *location* of an entity is a bounded geographic area with one or more “exits” at its border [4].

Definition 2 An *exit* is a necessary symbolic label through which two separated regions can be connected together, and it denotes a critical position where the location state is changing from one area to another. For example, a room exit is the place which marks people can leave “room” for “corridor” and vice versa.

Definition 3 A *location node* is a position element with several branches that point to its child nodes. The position elements to be used in our tree model include indoor exits, rooms, and corridor intersections.

Definition 4 The *root nodes* of the model tree will be the exits locating at the outer most boundary of the entire modeling space.

Definition 5 *Directly reachable* relation, denoted as $x \rightarrow y$, means that there exists a physical access (e.g., a corridor or stair for pedestrians or a road for automobiles) from location x to location y so that the path involves no other exits or rooms.

Definition 6 A *root path* of location x is a path from one root node to x without going through any other root nodes.

Definition 7 Location x is *under* location y , denoted as $x \setminus y$, if any root path of x goes through y , that is, if we reach x from a root exit node, y must be firstly passed before x .

Definition 8 Location x is *directly under* location y , denoted as $x \setminus y$, if $x \setminus y$ and no other location k satisfies $x \setminus k$ and $k \setminus y$ [4].

We can infer the following lemmas on the properties of the above “under” relations:

Reflexivity \forall location x , the root path of x must pass through itself, that is, $x \setminus x$;

Antisymmetry If there exist $x \setminus y$ and $x \neq y$, then $y \setminus x$ must be false.

Transitivity If $x \setminus y$ and $y \setminus k$, then k is in x ’s root path, that is, $x \setminus k$.

At the same time, “directly under” relation has the properties of irreflexivity and uniqueness, that is, there is one and only one y satisfies $x \setminus y$ for any x . If there is not any y satisfying $x \setminus y$, x must be a root. Also, if there exists $x \setminus y$, then $x \rightarrow y$ must be true.

“Directly under” relation-based location tree

In the combined method, we firstly model all the location nodes as a hierarchy so that human can easily understand and quickly index the hierarchy. And any one hierarchy requires a computable binary relation between elements, that is, the

answer to “do the two nodes have the key relationship?” would only be two cases: true or false.

The level classification in most existing hierarchical location models is based on “contained in” relation, so they cannot be developed into providing precise services for those applications based on the “connected to” relationship. Considering the advantages of hierarchical modeling and the higher requirements for navigation and the nearest queries, we choose the topological relation “directly under” as such a relation for hierarchical division. But the very first required condition is that our modeling situation has to meet the following two assumptions:

Assumption 1 All the indoor paths are two-way and their distances to the two ways are symmetric, that is, for the distance $\text{dist}(x, y)$ from location x to y , there must be

$$\text{dist}(x, y) = \text{dist}(y, x)$$

So, we just compute the distance only once and find out a best path P_b for one way, and then, the path for its opposite way could be created by arranging these nodes in the reverse order. For most interior layout, this hypothesis is reasonable.

Assumption 2 Any two exits located on either side of a common corridor must be directly reachable from each other, and the directly reachable path must also be unique and shorter than any other paths.

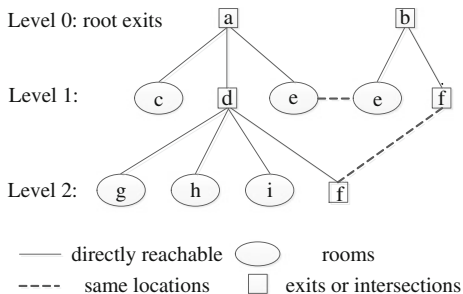
In the first step of modeling, it is necessary to gain the floor plan for a specific floor within a building or its inner layout. According to that, position elements are identified either manually or by program.

In the second step of modeling, our hierarchical model of locations $G = (V, E)$ will be constructed from the top to bottom based on the reachable semantic, “directly under” relation, both in graph and program. Here, V denotes a collection of location node vertices, and E is a collection of “directly under” edges. For any two location nodes x and y , if $x \setminus y$, then $\langle x, y \rangle \in E$. When a position element has multiple exits within different corridors, it may be repeated at different levels. Figure 1 illustrates such a location hierarchy model. All root exits are on the top of the hierarchy (level 0), other locations directly under to level 0, is in level 1, and so on. Level 0 vertices can denote building exits, while level 1 and 2 vertices can denote floor, corridor, or room, respectively. Formally, each vertex has a same single parent node except for the root exits and G is a tree structure because it suits for tree’s characteristics, such as without loop, connected graph and any nodes, except root node, can have none or more succeeding nodes.

The answer to “are the nodes at a same level directly reachable from each other?” will be given in definition 6:

Definition 9 *Level reachable* relation, denoted as $x \rightarrow\rightarrow y$, means that node x and node y are at the same level in G and they also have at least one common parent node.

Fig. 1 Example of a location hierarchy tree G



The level reachable relation has the following lemmas:

- (a) *Reflexivity* for \forall location x , $x \rightarrow\rightarrow x$;
Symmetry if $x \rightarrow\rightarrow y$, then $y \rightarrow\rightarrow x$.
Transitivity if $x \rightarrow\rightarrow y$ and $y \rightarrow\rightarrow k$, then x and k are also level reachable, that is, $x \rightarrow\rightarrow k$.
 If $x \rightarrow\rightarrow y$, then, $x \rightarrow y$.

3.2 Distance Identification Between Indoor Locations

For many location-based services, such as the nearest service discovery and shortest path planning, quantitative distances based on a predefined metric are required.

Definition 10 The distance from location x to location y is a *primitive distance*, denoted as $directLength(x, y)$, if x and y are directly reachable, that is, $x \rightarrow y$.

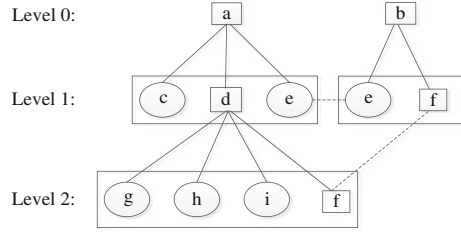
In our tree model, the primitive distances refer to the distances between all pairs of position elements which are “directly under” or “level reachable” from each other. Before providing services about accurate distances, the primitive distance must be preserved, so that it can be used as the weight for a related coordinate attributes and the edge of directly reachable relation. We can mark it manually or automatically by geometric information.

In the article, we generalize the **distance** semantics as the accumulation of all segment primitive distances on a specific path [11].

3.3 Set-Based Model

Based on the hierarchical tree model of locations shown in Fig. 1, several subsets of location coordinates can be created according to the “directly under” and “level reachable” topology relation defined above so as to provide indoor range querying.

Fig. 2 Example of core sets division



We call them “**core sets**,” denoted as S_c^i , because they are the bases for deriving other sets which correspond to different confined ranges. And in most cases, core sets are a collection of locations sharing a common corridor.

In general, a set of symbolic coordinates S forms the basis for the set-based approach. Different parts of modeled area can be defined by the subsets of universal set S . Now, let us consider a certain floor of a building, and Fig. 2 is its hierarchy graph. The set $S = \{a, b, c, d, e, f, g, h, i\}$ consists of all the position elements in this floor. The coordinates at level 1 in Fig. 2 can be described as two core sets, that is, $S_c^1 = \{a, c, d, e\}$ and $S_c^2 = \{b, e, f\}$, corresponding to two different corridors. Also, based on level 2, another core set $S_c^3 = \{d, g, h, i, f\}$ can be determined. In each core set, the first element is parent node of the succeeding leaf nodes.

It is easy to see that, the core sets based on hierarchy tree G can be used to calculate the partial overlapping locations when the intersection of two core sets S_c^i and S_c^j are not empty, that is $S_c^i \cap S_c^j \neq \emptyset$. However, any two core sets must not satisfy containment or subjection relation, that is, $S_c^i \cap S_c^j \neq S_c^i$ or $S_c^i \cap S_c^j \neq S_c^j$. On the other hand, according to G , if all the nodes in S_c^i are under a node in S_c^j , then $S_c^i \cap S_c^j \neq \emptyset$.

This set-based model can also be used to qualitatively compare the distances between symbolic coordinates by modeling the sets of neighboring symbolic coordinates. Here, we call this kind of set **neighborhoods**, denoted as S_{nei} . The set $S_{nei}(x, y)$ describes an area with the node x as the circle center and the distance from node x to node y as the radius of this circle. It is a collection of the following location coordinates: x, y and all other coordinates i satisfying $\text{dist}(x, i) < \text{dist}(x, y)$. By separating or merging the core sets around a specific point, we can derive any required neighborhoods corresponding to the different levels of maximum distance limitation away from that point. There is a lemma for explaining the relationship between distances and neighborhoods:

$$\text{for } \forall k, \exists x, y,$$

$$\text{dist}(x, y) < \text{dist}(x, k) \Leftrightarrow S_{nei}(x, y)_{nei}(x, k)$$

For instance, we suppose Fig. 3 shows a part of indoor room layout, the neighborhoods of coordinate b related to a, c and e are, respectively, as follows:

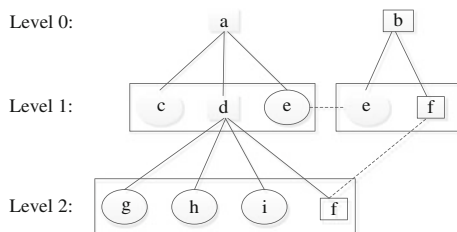


Fig. 3 Example of neighborhoods for two locations

$$\begin{aligned}
 S_1 &= S_{\text{nei}}(b, a) = \{a, b\}; \\
 S_2 &= S_{\text{nei}}(b, c) = \{a, b, c\} \\
 S_3 &= S_{\text{nei}}(b, e) = \{a, b, c, d, e\}
 \end{aligned}$$

It is thus clear that the three sets S_1, S_2 and S_3 satisfied $S_1 \subset S_2 \subset S_3$, so we can come to the following conclusion:

$$\text{dist}(b, a) < \text{dist}(b, c) < \text{dist}(b, e)$$

In practice, this qualitative distance lemma can sometimes be used to substitute for quantitative distances computing and effectively compare distances between a pair of locations, because qualitative calculations have relatively lower cost.

3.4 A Path Planning Algorithm

Based on the combined hierarchy-set model and the primitive distances, we propose a new procedure for the shortest path search and navigation, which can be programmatically constructed. The below algorithm 1 illustrated this procedure in detail.

In a smart space, we can sense the abnormal environment information in real time. With more than one candidate paths, users would be provided much more suitable and personalized suggestions or services. That is the primary purpose for context-aware application systems to achieve.

The major difference between former models and ours lies in that, in our model, the symbolic hierarchies and core sets are constructed in accordance with the geometric attributes, that is, based on the hierarchical tree, any “directly under”

Algorithm 1 Find the best path P_b in the combined model

Input:

G (“directly under”-based layer division);

directLength (primitive distances);

S_c^i (“level reachable”-based region division);

x, y (the source and destination node)

(continued)

(continued)

Algorithm 1 Find the best path P_b in the combined model

Procedure:

Specific core sets' identification:

query and identify the homologous core sets S_{cx} and S_{cy} , which contain x and y as its leaf node, respectively. If an input node is an intersection, that is, more than one core sets contain it, then S_{cx} or S_{cy} should be considered as a set of core sets.

Directly reachable path:

if $\exists y \in S_{cx}^i$ or $x \in S_{cy}^j$, that is, there is a core set containing both location x and location y . At the moment, x and y are obviously directly reachable from each other resulting from $x \rightarrow y$ or $x \setminus y$ or $y \setminus x$, then P_b is the sole directly reachable path $x \rightarrow y$ (according to assumption 2), break;

Candidate paths planning:

else if S_{cx} and S_{cy} have one or more intersection nodes but neither x nor y , that is, $S_{cx}^i \cap S_{cy}^j = I \neq \emptyset$, and $\forall i \in I$ must satisfy $i \neq x$ and $i \neq y$, **then**, save $x \rightarrow i \rightarrow y$ as a destination path P_i (Fig. 4 has identified the candidate path from node c to f , the planning path resulting from this step is $c \rightarrow e \rightarrow f$);

else if there exists a core set S_{cc}^k in which at least one leaf node is under two nodes, which are from S_{cx} and S_{cy} respectively (non-eliminating the special case $S_{cc}^k = S_{cy}^j$), **then** start from x , track for y form bottom to top along with the “direct under” branches in G , and locate all the tree-based paths P_{down} in terms of $x \rightarrow s_{cx}^{ij} \rightarrow s_{c(x+1)} \rightarrow \dots \rightarrow s_{cc} \rightarrow \dots \rightarrow s_{c(y+1)} \rightarrow s_{cy}^{ij} \rightarrow y$ (corresponding to $c \rightarrow d \rightarrow f$ in Fig. 4);

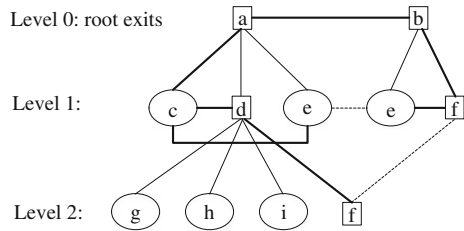
else start from x , track for y from top to bottom along with the branches in G , locate all the tree path P_{up} in terms of $x \rightarrow s_{c(x-1)} \rightarrow \dots \rightarrow s_{cc} \rightarrow \dots \rightarrow s_{c(y-1)} \rightarrow y$ (corresponding to $c \rightarrow a \rightarrow b \rightarrow f$ in Fig. 4);

Best path selection:

separately calculate the distances of P_i, P_{down}, P_{up} by adding up the primary distance *directLength* for each segment in a path. And sort these paths in ascending order according to their distances, then save them as an array P_b ;

represent the suggested path. In array P_b , the first element is the shortest path between nodes x and y . While in some situations, the shortest path may not be the most satisfying answer because of something like traffic jams, then another path element of P_b can be provided.

Fig. 4 Example of candidate path planning



relation branch connecting location x and y can qualitatively default as the best path between x and y which are reachable from each other. That will exclude many unnecessary paths and greatly reduce computing cost for path planning. On the other hand, once the model is built, high-level location navigation, browsing, and search can be implemented automatically through the well-defined topological relations and semantic distance.

4 A Complete Example: Modeling an Indoor School Hospital

In this section, we illustrate the whole modeling process toward a specific physical area. It is a portion of the university hospital in Beijing Institute of Technology (BIT)

Step 1 We need to depict the floor plan of the modeling area shown in Fig. 5 and identify all the position elements to be modeled manually or programmatically by computer graphics technology.

In this example, the following geographic areas are identified as position elements: Room 0101 ~ 1004, corridor intersections 01 ~ 10, and building exit *a*, *b* and *c*.

Step 2 According to the principle of location hierarchical tree *G*, here, we describe it in a more intuitive graph (Fig. 6). It is necessary to note that, for further automatic distance-related LBS, such as path planning, the nearest printer querying, and so on, *G* should be implemented programmatically, and the primitive distances should also be retrieved and stored in our model *G* as the edge weights. Exit *a*, *b* and *c* are chosen as the root exits and that means they are the only accesses for leaving the modeled space.

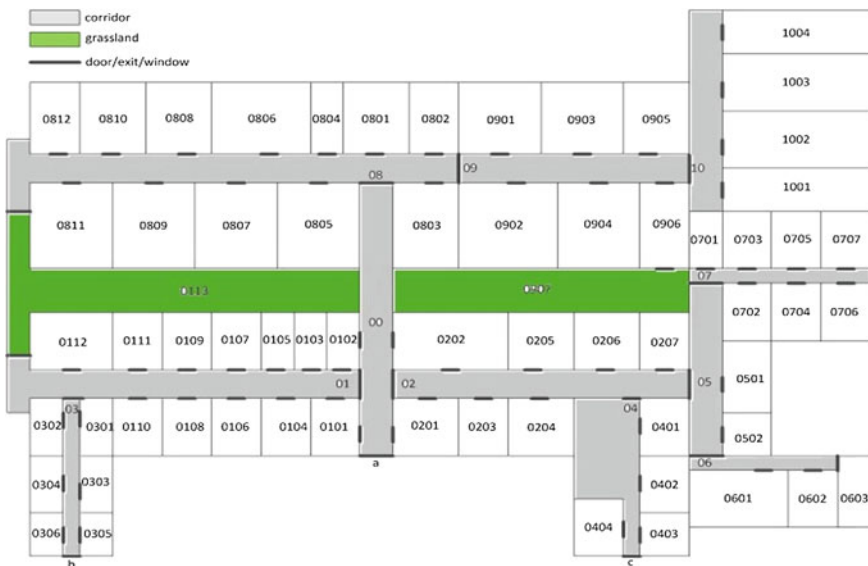


Fig. 5 Floor plan of BIT school hospital

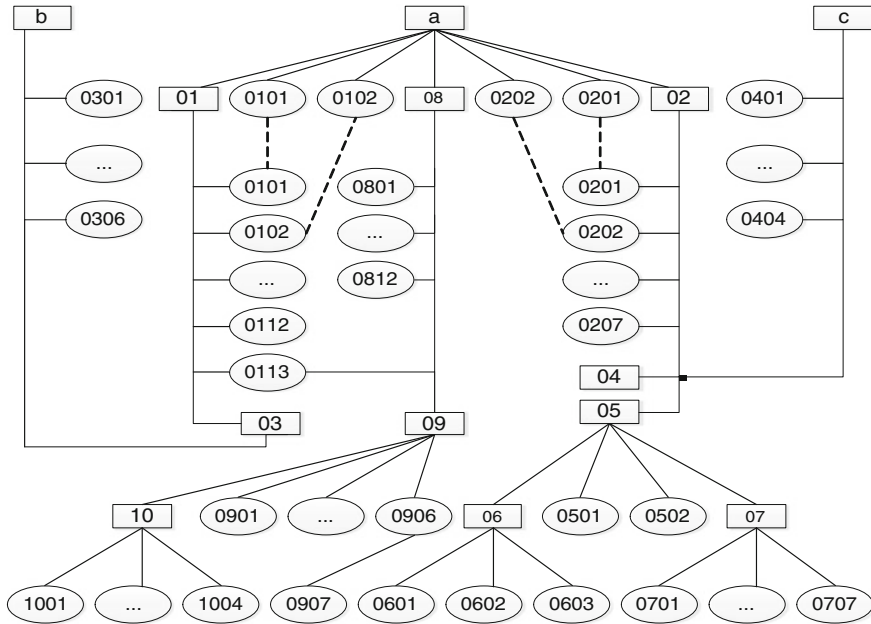


Fig. 6 Hierarchy tree model G of locations in the hospital

Step 3 By traversing the location hierarchy tree presented in step 2 and classifying the nodes having a common parent node together, we can programmatically create all the core sets in G . The resultant core sets for the hospital are as follows:

$$\begin{aligned}
 S_c^{00} &= \{a, 01, 02, 08, 0101, 0102, 0201, 0202\}; \\
 S_c^{01} &= \{01, 03, 0101, 0102, 0103, \dots, 0113\}; \\
 S_c^{02} &= \{02, 04, 05, 0201, 0202, \dots, 0207\}; \\
 S_c^{03} &= \{b, 03, 0301, 0302, \dots, 0306\}; \\
 S_c^{04} &= \{c, 04, 0401, 0402, 0403, 0404\}; \\
 S_c^{05} &= \{05, 06, 07, 0501, 0502\}; \\
 S_c^{06} &= \{06, 0601, 0602, 0603\}; \\
 S_c^{07} &= \{07, 0701, 0702, \dots, 0707\}; \\
 S_c^{08} &= \{08, 09, 0801, 0802, \dots, 0812, 0813\}; \\
 S_c^{09} &= \{09, 10, 0901, 0902, \dots, 0906\}; \\
 S_c^{10} &= \{10, 1001, 1002, 1003, 1004\}; \\
 S_c^{11} &= \{0906, 0907\}
 \end{aligned}$$

5 Conclusion and Future Work

In this paper, a combined semantic location modeling method for indoor layouts was proposed. The proposed model consists mainly of two parts: the “directly under”-based hierarchy tree of locations and the “level reachable”-based core sets of symbolic coordinates. Based on a series of notion definitions and lemmas, we generated a hierarchy tree model which retrieves all primitive segment distances as the edge weight and integrates various position elements into a single model. Then, we extended it by classifying and creating core sets in order for further area identifying and easier path planning programmatically. Thus, all the basic semantics, including “reachable” relation, “contained in” relation and “distance,” are abstracted, stored, and then seamlessly integrated into this model. At last, by making full use of the core sets and the advantage of the hierarchy model in indexing, we proposed an efficient and convenient path searching algorithm.

Our model not only can be used to provide navigation and the nearest query for mobile users according to the location tree and accumulative path distance, but also can be used to range query and further promote the user’s current location context to be concisely and unambiguously displayed on his handheld device. In a word, our model can balance more functions well, such as path planning and entity browsing, and with less cost.

As the future work, we are to integrate the location model into an ontology-based context model by mapping room coordinates to the related business or event context, considering delayering system model at the same time. Based on the models, we will develop a prototype system involving location querying, context reasoning, activities guiding, and workflow to improve the agility and reliability of business process.

References

1. Weiser M (1991) The computer for the 21st century. *Sci Am* 265:94–104
2. Schilit B, Adams N, Want R (1994) Context-aware computing applications. In: Proceedings of the IEEE workshop on mobile computing system and application, pp 85–90
3. Abowd G, Dey A, Brown P, Davies N, Smith M, Steggle P (1999) Towards a better understanding of context and context-awareness. In: Gellersen HW (ed) *Handheld and ubiquitous computing*. Springer, Berlin, pp 304–307
4. Hu H, Lee DL (2004) Semantic location modeling for location navigation in mobile environment. In: Proceedings of the 2004 IEEE international conference on mobile data management (MDM’04), pp 52–61
5. Becker C, Dürr F (2005) On location models for ubiquitous computing. *Pers Ubiquit Comput* 9:20–31
6. Bettini C, Brdiczka O, Henriksen K (2010) A survey of context modelling and reasoning techniques. *Pervasive Mob Comput* 6:161–180
7. Ward A, Jones A, Hopper A (1997) A new location technique for the active office. *Pers Commun, IEEE* 4:42–47
8. Want R (1992) The active badge location system. *ACM Trans Info Syst (TOIS)* 10:91–102

9. Kidd C (1999) “The aware home: A living laboratory for ubiquitous computing research,” Cooperative buildings. Integrating information, organizations, and architecture, pp 191–198
10. Bauer M, Becker C, Rothermel K (2002) Location models from the perspective of context-aware applications and mobile ad hoc networks. *Pers Ubiquit Comput* 6:322–328
11. Zhang Z, Yang J, Gu J, Xu Y (2009) Event based semantic location model in cooperative mobile computing. pp 203–208

Acquisition of Class Attributes Based on Chinese Online Encyclopedia and Association Rule Mining

Zhen Jia, Hongfeng Yin and Dake He

Abstract A new method of extracting class attributes based on Chinese online encyclopedia and association rules mining is proposed in this paper. Unstructured texts of encyclopedia articles are taken as the extracting source. Attribute values are viewed as named entities and the frequent pattern mining technique is applied in the proposed method. The candidate attribute words are generated through mining association rules between words and named entities in frequent patterns. Synonymous or duplicate candidate attribute words are found based on the external semantic resource and the computation of similarity of words. Uniform attribute names are obtained by filtering synonymous candidate attribute words. Unstructured texts of six classes are downloaded from Chinese online encyclopedia as experimental data. The experimental results indicate the proposed method is feasible and effective.

Keywords Knowledge harvesting · Information extraction · Class attribute extraction · Chinese online encyclopedia · Frequent pattern mining

1 Introduction

Attribute is characteristic of the different classes of things. For example, the attributes of person will be name, birth date, birth place, occupation, and so on. Acquisition of knowledge about attributes includes acquisition of class attributes

Z. Jia (✉) · H. Yin · D. He
School of Information and Science Technology, Southwest Jiaotong University,
610031 Chengdu, China
e-mail: zjia@home.swjtu.edu.cn

and attribute values of instances for given classes. Online encyclopedias, such as Wikipedia, Baidu Baike, and Hudong Baike, are created by network users collaboratively. Online encyclopedia includes a large number of classes and instances which contain abundant knowledge about attributes. Extracting knowledge about attributes from online encyclopedia is of great significance for building large-scale knowledge base or automatic question and answering system [1]. To extract knowledge about attributes from Chinese online encyclopedia, we need to acquire class attributes and give the uniform attributes names for each class. Many large-scale universal knowledge bases (e.g., DBpedia [2], Yago [3]) take Wikipedia as one of the knowledge sources and make use of Wikipedia's features to obtain facts information. For example, infoboxes contain many valuable class attributes and attributes value information of instances. However, infoboxes suffer from several problems. Schema drift [4] is one of them. The reason is because users are free to create or modify infoboxes templates and the infoboxes schema tends to evolve during the course of authoring [4]. It is necessary to redefine infoboxes templates and give uniform attribute names for each class. For Chinese online encyclopedia (e.g., Hudong Baike and Baidu Baike), only some of entries have infoboxes. Baidu Baike infoboxes templates are created by users and have the same problems as Wikipedia. Infoboxes of Hudong Baike have uniform templates for some classes. However the attributes are common attributes of the classes. For the above reasons, we exploit unstructured texts of Chinese online encyclopedia articles as extracting source and apply frequent patterns mining technique to acquire the class attributes from the texts. There have some related works on class attribute extraction. Tokunaga et al. [5] proposed an unsupervised method of acquiring attribute words from Web documents that utilize statistics on words, lexico-syntactic patterns, and HTML tags. Because there are no standard definitions of attributes, or criteria for evaluating obtained attributes, they proposed criteria of question-answerability to evaluate the attribute words. Pasca et al. [6] took the logs of Web search queries as the extracting source and proposed a method to extract class attributes which consists of three steps: selection of candidate attributes for the given set of classes; filtering of the attributes for higher quality; and ranking of the attributes. However, this method relies heavily on query logs. Pasca et al. [7] proposed another method to extract class attributes from a combination of both documents and search query logs. Kopluku et al. [8] use HTML tables to extract and rank attributes.

The organization of the rest of the paper is as follows. In Sect. 2, we review the necessary background of association rule problem and frequent patterns mining. In Sect. 3, we present the process of extracting class attributes. In Sect. 4, experiment results are discussed. We conclude with a summary in Sect. 5.

2 Background

Association rules mining is one of the main research areas of data mining. It aims at finding the interesting relations from a large number of itemsets or patterns. The basic problems of association rules are described as follows [9].

Let $I = \{i_1, i_2, \dots, i_n\}$ be a set of all items. Let $D = \{t_1, t_2, \dots, t_m\}$ be a set of transactions. Each transaction in D has a unique transaction ID and contains a subset of the items in I . A rule is defined as an implication of the form $A \rightarrow B$ where $A, B \subseteq I, A \cap B = \emptyset$. A and B are called antecedent and consequent of the rule, respectively.

To select interesting rules from the set of all possible rules, constraints on various measures of significance and interest can be used. The best-known constraints are minimum thresholds on support and confidence. The support of A is defined as $\text{supp}(A)$. $\text{Supp}(A)$ is the frequency of transactions in the data set which contain A . The confidence of a rule $A \rightarrow B$ is defined as $\text{conf}(A \rightarrow B) = \text{supp}(A \cup B) / \text{supp}(A)$. Support reflects the importance of rules. Confidence reflects the dependency relations between antecedent and consequent of rules. Association rules are usually required to satisfy a user-specified minimum support and a user-specified minimum confidence at the same time.

Association rule generation is usually split up into two separate steps. First, minimum support is applied to find all frequent itemsets in a database. Second, these frequent itemsets and the minimum confidence constraint are used to form rules.

3 Class Attributes Extraction Method

3.1 Process of Class Attribute Extraction

In our method, attribute values are viewed as named entities [10]. Our method is based on the observation found that attribute values tend to co-occur with attribute words within a few words in the sentences. If some words or phrases tend to co-occur with named entities, these words or phrases will have relationship with the named entities and may be candidate attribute words. The process of extracting class attributes consists of the following four steps.

- Frequent patterns mining
- Association analysis
- Extraction of candidate attribute words
- Acquisition of uniform attribute names

3.2 Frequent Patterns Mining

In the step of frequent patterns mining, we extract k words sequences from texts for a given class. K can be equal to 1, 2, Different from n -grams patterns extraction [11], the words in our patterns can be non-consecutive. In the patterns, if there are words which are named entities, the words are replaced by their named entity tags.

For example, in a 2-tuple pattern (校长, nr), nr is the tag of person name. The word 校长 means *headmaster*. We apply technique of frequent patterns mining to find frequent k -tuple patterns. Due to the greater the distance between the words is in the sentence, the looser the relationship between them is, we use the window to restrict the range of extraction.

Let N be the number of words in a window. The number of words we can extract in a window is no greater than N . The moving of the window is performed word by word.

After word segmentation, POS tagging, and named entity tagging, a sentence can be viewed as a sequence of words and tags. A sentence is symbolized by a sequence $S = \{(w_1, p_1), (w_2, p_2), \dots, (w_L, p_L)\}$.

The lower case letter w stands for word, p stands for POS tag or named entity tag, and the subscript number stands for the position of word in the sentence.

For example, a sentence consists of six words and the third word in the sentence is a named entity. This sentence is symbolized by the sequence $S = \{(w_1, p_1), (w_2, p_2), (w_3, p_3), (w_4, p_4), (w_5, p_5), (w_6, p_6)\}$. If N is equal to 4 and the window is in the first word of the sentence, extracted k -tuple patterns are as follows:

1-tuple patterns $\{(w_1)\}$
 2-tuple patterns $\{(w_1, w_2), (w_1, p_3), (w_1, w_4)\}$
 3-tuple patterns $\{(w_1, w_2, p_3), (w_1, w_2, w_4), (w_1, p_3, w_4)\}$
 4-tuple patterns $\{(w_1, w_2, p_3, w_4)\}$

The third word is replaced by its named entity tag p_3 . With sliding of the window, we extract k -tuple patterns ($k = 1, 2, 3, 4$) until the end of the sentence. In order to avoid extracting duplicate patterns, we only extract the patterns including the first word in the window. For example, when the window moves to the second word of the sentence, extracted patterns are as follows:

1-tuple patterns $\{(w_2)\}$
 2-tuple patterns $\{(w_2, p_3), (w_2, w_4), (w_2, w_5)\}$
 3-tuple patterns $\{(w_2, p_3, w_4), (w_2, p_3, w_5), (w_2, w_4, w_5)\}$
 4-tuple patterns $\{(w_2, p_3, w_4, w_5)\}$

When the window moves to the end of the sentence or the number of words of sentence is less than N , we treat the sentence as a transaction and extract all k -tuple patterns in order not to the omission of any patterns. For reducing quantity of extracted patterns and ignoring useless patterns, we only extract the words which belong to certain lexical categories (e.g., verb, noun, adjective).

The support of a pattern is defined as $\text{supp}(\text{pattern})$. $\text{Supp}(\text{pattern})$ is the number of times the pattern appears. If $\text{supp}(\text{pattern})$ is no less than the minimum support, the pattern is a frequent pattern.

3.3 Association Analysis

If a frequent pattern consists of words and named entity tags, we compute the confidence between the words and named entity tags. For example, (耕地, 面积, mq) is a 3-tuple pattern in which mq is a quantifier named entity tag and (耕地, 面积) is a words sequence. Here 耕地 means *cultivated land* and 面积 means *acreage*. The confidence of an association rule (耕地, 面积) \rightarrow mq is equal to $\text{supp}(\text{耕地, 面积, } mq) / \text{supp}(\text{耕地, 面积})$. (耕地, 面积) is a 2-tuple pattern. If confidence of the rule is no less than the minimum confidence, pattern (耕地, 面积, mq) becomes a candidate attribute pattern.

According to the types of named entity tags in patterns, the patterns are categorized into several types: quantity patterns, location patterns, person patterns, time patterns, and so on. For different types of patterns, we specify different minimum support and confidence to mine association rules.

Some candidate attribute patterns are subsequences of other candidate patterns. For example, pattern (耕地, mq) is subsequence of pattern (耕地, 面积, mq). In this situation, the short patterns are deleted because the meanings that long patterns express are more specific.

3.4 Extraction of Candidate Attribute Words

The words in candidate attribute patterns are extracted as candidate attribute words. For example, in candidate pattern (耕地, 面积, mq), word sequence (耕地, 面积) are candidate attribute words. mq is the type of attribute values. In candidate pattern (校长, nr), (校长) is a candidate attribute word. nr is the type of attribute values.

In Table 1, we list ten more examples of candidate attribute words extracted from candidate attribute patterns for town class. (We added possible English translations for the attribute words)

3.5 Acquisition of Uniform Attribute Names

There are some candidate attribute words semantically close to one another. For example, in Table 1, “位于” and “地处” are synonymous words and the meaning

Table 1 Candidate attribute words of town class

Candidate pattern	Candidate attribute word		Attribute value
(辖, <i>mq</i> , 行政村)	辖, 行政村	Jurisdiction, village	<i>mq</i>
(下辖, <i>mq</i> , 行政村)	下辖, 行政村	Jurisdiction, village	<i>mq</i>
(幅员, 面积, <i>mq</i>)	幅员, 面积	Land, area	<i>mq</i>
(国土, 面积, <i>mq</i>)	国土, 面积	Land, area	<i>mq</i>
(位于, <i>ns</i>)	位于	Locate	<i>ns</i>
(地处, <i>ns</i>)	地处	Locate	<i>ns</i>
(年, 平均, 气温, <i>mq</i>)	年, 平均, 气温	Annual, average, temperature	<i>mq</i>
(人均, 耕地, <i>mq</i>)	人均, 耕地	Per capita, arable land	<i>mq</i>
(最高, 海拔, <i>mq</i>)	最高, 海拔	Highest, height above sea level	<i>mq</i>
(总, 人口, <i>mq</i>)	总, 人口	Total, population	<i>mq</i>
(农业, 总产值, <i>mq</i>)	农业, 总产值	Agriculture, total value of out-put	<i>mq</i>

of them is *locate* in English. Synonymous words lead to duplicate attributes for a given class.

To filter the synonymous words, we firstly use synonyms dictionary [12] to find synonymous words pairs of candidate attributes and replace synonymous words in patterns with standard words. Standard words are words that have higher support in synonymous words pairs. For example, “地处” is replaced by “位于” in patterns because the number of times “位于” occurs is more than “地处” in the texts of town class. We replace “幅员” with “国土” because the number of times “国土” occurs is more than “幅员” in the texts of town class. Secondly, we compute similarity of the candidate attribute words based on their Jaccard similarity coefficient. For example, the similarity of “辖, 行政村” and “下辖行政村” is equal to 80%. If the similarity of two attribute words meets the similarity threshold, the attribute words with less support will be filtered. By using external semantic resource and computing the similarity of attribute words, we filter synonymous or duplicate words and obtain uniform attribute words.

4 Experimental Results

4.1 Data Set

The experiments were performed on six classes of Hudong Baike as of November 2011. The six classes are university, town, village, company, factory, and middle school. To find entry names for each class, we adopt several approaches such as category tags clustering [13], literal similarity of entry names [14], and Sogou dictionaries [15]. The data of six classes we downloaded were about 60,000 articles. We apply SWJTU Chinese Word Segmentation System [16] to preprocess the data.

4.2 Experiment on Extraction of Patterns

We extract k -tuple patterns from the texts of each class. When N is equal to 4, the number of different k -tuple patterns ($k = 1, 2, 3, 4$) is shown in the Fig. 1.

For each class, the number of different 3-tuple patterns is larger than 4-tuple patterns and 2-tuple patterns. When we change the value of N , the number of different patterns also changes. For example, in town class, the number of different patterns changing with N is shown in Fig. 2. When N increases, the number of different patterns increases, too.

According to the named entity tags in the patterns, the patterns are classified into several types that are quantity patterns, location patterns, person patterns, time patterns, crop patterns, and other patterns. Fig. 3 shows that in the texts of six classes, the number of quantity patterns is the largest, followed by location patterns, time patterns, person patterns, and crop patterns. There are a few of crop patterns in the texts of town, village, and company classes. We do not analyze other types of patterns because other types are so few in number.

Fig. 1 Statics of extracted patterns when N is equal to 4

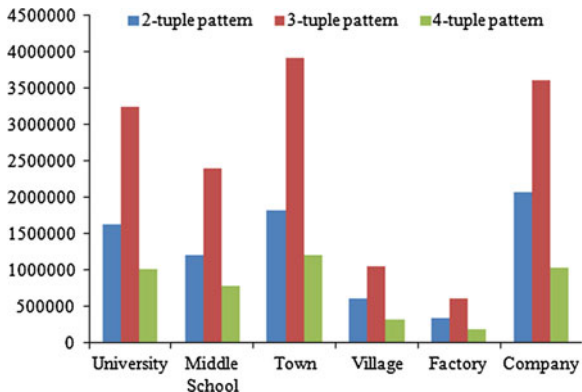
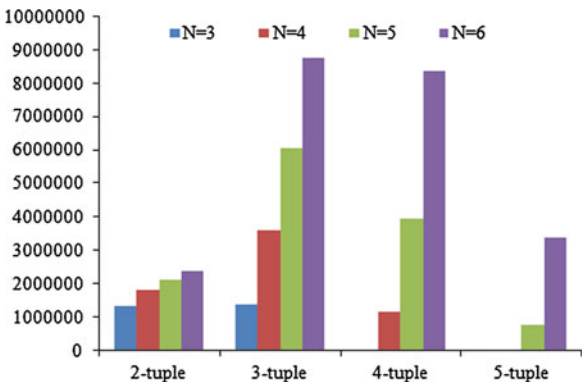


Fig. 2 The number of patterns changes with N for town class



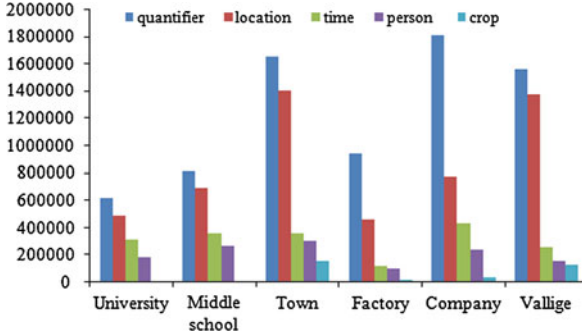


Fig. 3 The number of patterns changes with N for town class

4.3 Experiment on Association Analysis

According to the different types of patterns, we specify the minimum support and confidence threshold, respectively, to find interesting patterns. When value of N changes, the number of interesting patterns will change, too. Table 2 lists the number of interesting patterns which will be candidate attribute patterns for town class.

In candidate attribute words, there are some patterns that are not interesting. For example, the confidence of (先生) $\rightarrow nr$ is very high because the word “先生” often co-occurs with nr named entity tag (“先生” means *Sir*). But (nr , 先生) is not an interesting pattern. Candidate attributes patterns that are not interesting have to be removed manually. If the minimum support or confidence is specified too low, more candidate attribute patterns are acquired, but there are more synonymous or uninteresting patterns.

Besides the minimum support and confidence, the value of N also affects the number of candidate attribute patterns. When the value of N changes from 3 to 6, given the minimum support and confidence, the number of candidate attribute patterns will change with N .

Table 2 The number of interesting patterns for town class

Type of pattern	min_supp	min_conf	$N = 3$	$N = 4$	$N = 5$	$N = 6$
Quantity	150	0.7	133	160	323	456
Location	500	0.6	3	13	33	46
Person	50	0.3	10	37	56	134
Time	80	0.8	3	32	61	65
Crop	50	0.6	7	24	41	67
Total			158	257	657	762

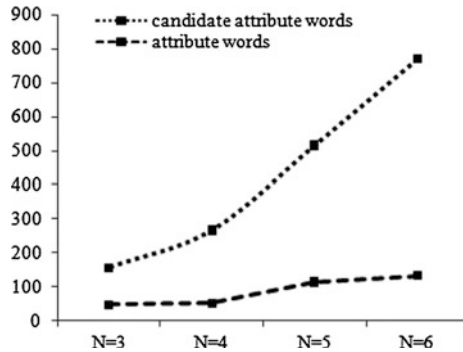


Fig. 4 Attribute words change with N

Table 3 Number of extracted attributes over target classes

Class	Extracted attributes					Total
	Quantity	Location	Person	Time	Others	
University	55	1	6	1	0	63
Middle school	43	1	2	1	0	47
Town	113	6	4	4	4	131
Village	176	2	5	3	5	191
Factory	21	4	5	1	1	32
Company	19	7	10	1	1	38

4.4 Experiment on Acquisition of Uniform Attribute Names

After filtering of synonymous and similar words in candidate attribute words, attributes for a given class are acquired.

Figure 4 shows candidate attribute words, and acquired attributes changes with N for the town class. From Fig. 4, we can see the number of candidate attribute words has a rapid increase with increase of N. However, more synonyms, wrong, meaningless, and uninteresting candidate attribute words appear with increasing of N, so acquired attributes do not increase rapidly. That is because the greater N is the greater the distance of the words in patterns is and the number of meaningless patterns increase. The number of acquired attributes and some attributes examples for six target classes are shown in Appendix.

Table 4 Examples of attributes for target classes

Class	Attributes
University	地址, 建校年代, 学生人数, 教职工, 纸质图书, 本科学制, 专职教师, 占地面积, 总建筑面积, 校长 Location, founded time, student number, staff number, collected books number, schooling length, full-time teacher number, floor space, overall floorage, headmaster
中学	地址, 占地面积, 始建于, 教职工, 联系人, 现有教学班, 一级教师, 校长, 二级教师, 建筑总面积 Location, floor space, founded time, staff number, contact person, class number, first-grade teacher number, headmaster, second-grade teacher number, overall floorage
Middle school	Location, floor space, founded time, staff number, contact person, class number, first-grade teacher number, headmaster, second-grade teacher number, overall floorage
乡镇	位于, 总面积, 总人口, 农民人均纯收入, 下辖行政村, 总户数, 主产, 非农业人口, 农业总产值, 人均耕地面积 Location, total area, total population, rural per capita net income, administrative villages number, total households, main product, non-agricultural population, total agricultural output value, per capita actual area of cultivated land
Town	Location, total area, total population, rural per capita net income, administrative villages number, total households, main product, non-agricultural population, total agricultural output value, per capita actual area of cultivated land
村庄	位于, 隶属于, 人均耕地, 外出务工, 主要种植, 农户数, 年平均气温, 农田面积, 农村经济总收入, 种植业收入 Location, under the jurisdiction of, per capita actual area of cultivated land, number of going out for non-farming job, mainly plant, total households, annual average temperature, farmland area, total income of the rural economy, income from plant industry
Village	Location, under the jurisdiction of, per capita actual area of cultivated land, number of going out for non-farming job, mainly plant, total households, annual average temperature, farmland area, total income of the rural economy, income from plant industry
工厂	员工人数, 法定代表人, 地址, 成立时间, 年营业额, 厂房面积, 月产量, 年出口额, 注册资本, 联系人 Employee number, legal representative, address, founded time, annual sales volume, factory area, monthly output, annual volume of export, registered capital, contact person
Factory	Employee number, legal representative, address, founded time, annual sales volume, factory area, monthly output, annual volume of export, registered capital, contact person
公司	销售收入, 公司总部位于, 董事长, 成立日期, 注册地址, 集团总裁, 员工人数, 股票代码, 营业收入, 开户行 Sales revenue, corporate headquarters location, chairman, founded time, registered address, group CEO, employee number, stock code, business income, opening bank
Company	Sales revenue, corporate headquarters location, chairman, founded time, registered address, group CEO, employee number, stock code, business income, opening bank

5 Conclusion

This paper proposed a method of extracting class attributes based on the observation that attributes values co-occur with attribute words within a few words in the sentences. We take the unstructured texts from Hudong Baike, Chinese online encyclopedia, as extracting source because online encyclopedia has rich knowledge about attributes of instances for many classes.

In this paper, frequent patterns and association rules mining techniques are applied to find candidate attribute patterns. Since users are free to create or modify encyclopedia articles, many attribute words are synonymous or semantically close to one another. We use synonyms dictionary to find synonymous attributes and compute the similarity of words to filter attribute words for higher quality. However, because of syntax errors in texts and imperfection in the course of

natural language preprocessing, wrong, uninteresting, and meaningless attributes need to be removed manually. The experiment results indicate this method is effective. Besides the six classes mentioned in the paper, this method can be used to acquire attributes of other classes and scaled to other corpus for a given class. While the proposed method relies on named entity recognition, the texts need to be classified.

Our contribution in the paper is an important stage of constructing large-scale facts knowledge base. In the next stage, we will identify the values of the attributes for various instances of the given classes.

Acknowledgments This work was supported by the National Natural Science Foundation of China Director Fund No.61152001 and the Open Project of Lab. of Management and Control for Complex Systems, Chinese Academy of Sciences No. 20110102.

Appendix: Number of Extracted Attributes and Examples of Attributes Over Target Classes

Table 3 lists the number of extracted attributes for the six target classes. According to the types of attribute value, attributes are categorized into quantity, location, person, time, and other classes. Table 4 lists some examples of attributes for the six target classes.

References

1. Jun Z, Kang L, Guangyou Z, Cai L (2011) Open information extraction. *J Chin Inf Process* 25:98–110
2. Auer S, Bizer C, Lehmann J, Kobilarov G, Cyganiak R, Ives Z (2007) DBpedia: a nucleus for a web of open data. 6th International Semantic Web Conference, Springer, Berlin
3. Suchanek FM, Kasneci G, Weikum G (2007) Yago: a core of semantic knowledge-unifying WordNet and Wikipedia, in Proceedings of the 16th international conference on, World Wide Web
4. Fei W, Daniel SW (2007) Autonomously semantifying Wikipedia. In: Proceedings of CIKM, pp 41–50
5. Tokunaga K, Kazama J, Torisawa K (2005) Automatic discovery of attribute words from web documents. In: Proceedings of the second international joint conference on natural language processing, pp 106–118
6. Pasca M, Durme BJ (2007) What you seek is what you get: extraction of class attributes from query logs. In: Proceedings of international joint conferences on, artificial intelligence, pp 2832–2837
7. Pasca M, Durme BJ (2008) Weakly-supervised acquisition of open-domain classes and class attributes from web documents and query logs. In: Proceedings of association for, computational linguistics, pp 19–27

8. Kopliku A, Sauvagnat K, Boughanem M (2011) Retrieving attributes using web tables. In: Proceedings of the 11th annual international ACM/IEEE joint conference on digital libraries, pp 13–17
9. Agrawal R, Srikant R (1994) Fast algorithm for mining association rules. In: Proceedings of the 1994 International conference on very large data bases. Santiago, Chile
10. Satoshi S (2008) Extended named entity ontology with attribute information. In: Proceedings of the sixth international language resources and evaluation
11. Nakashole N, Weikum G, Suchanek F (2012) PATTY: a taxonomy of relational patterns with semantic types. In: Proceedings of EMNLP
12. Wanxiang C, Zhenghua L, Ting L (2010) LTP: a Chinese language technology platform. In: Proceedings of the Coling (2010) demonstrations. Beijing, China, pp 13–16
13. Zhen J, Hongfeng Y, Tianrui L Research on Chinese online encyclopedia open category hierarchy tree and clustering Algorithms. Application Research of Computers
14. Xueying Z, Guonian L (2008) Approach to conversion of geographic information classification schemes. J remote sens 12:433–440
15. Sogou Dictionary, <http://pinyin.sogou.com/dict/>
16. Chinese Word Segmentation, <http://www.yebol.com.cn/>

A Delivery Time Computational Model for Pervasive Computing in Satellite Networks

Jianzhou Chen, Lixiang Liu and Xiaohui Hu

Abstract For developing the ubiquitous infrastructures of pervasive computing, low earth orbit (LEO) satellite networks, with the ability to provide global broadband access service, are considered as the complement and extension of the terrestrial networks. Due to the dynamic topology and the non-uniform distribution of terrestrial users, delivery time within LEO satellite networks is liable to fluctuate, which might deteriorate the capacity of accessing computing resource for users. Thus, to investigate the time constraint, a delivery time computational model for LEO satellite networks is proposed. With source inputs of Markov-modulated Poisson process (MMPP) and a hot-spot distribution of destinations, a tandem queue for the target flow is established with cross-flows. Then departure interval moments of the target flow are calculated and fitted for the next link as input parameters, from which the queuing delay under cross-flows is iteratively obtained. The comparison between computational results and simulation results demonstrates that this model makes a good depiction of the influence by the traffic pattern with satisfying accuracy.

Keywords Satellite networks · Delivery time · MMPP · Tandem queue

J. Chen (✉) · L. Liu · X. Hu
Science and Technology on Integrated Information System Laboratory,
Institute of Software, Chinese Academy of Sciences, Beijing, China
e-mail: chenjianzhou1986@126.com

L. Liu
e-mail: lixiang@iscas.ac.cn

X. Hu
e-mail: hxx@iscas.ac.cn

J. Chen
University of Chinese Academy of Sciences, Beijing, China

1 Introduction

With notable progress in computing capability of mobile and embedded devices, as well as the emerging support for cloud computing and social network, pervasive computing is facing a bright prospect, which would create ambient intelligence in our physical surroundings. However, the promising blueprint could be hope in vain without continual and reliable connectivity, which saddles ubiquitous infrastructures with great responsibility. Since typical terrestrial wireless access infrastructures such as Global System For Mobile Communication (GSM), Universal Mobile Telecommunications System (UMTS), and Wireless Local Area Network (WLAN) that are scant in rural areas could not guarantee seamless coverage, satellite networks, as the complement and extension of terrestrial networks, shed new light for pervasive computing. Particularly, low earth orbit (LEO) satellite networks with moderate propagation delay and low terminal power requirement have been regarded as a potential solution to provide global broadband access services, especially after the emergence of onboard processing capability and gigabit links [1]. But, due to the dynamic topology and the non-uniform distribution of terrestrial traffic intensity, delivery time within LEO satellite networks is liable to fluctuate, which might deteriorate the capacity of accessing computing resource for users. Meanwhile, as the demand of real-time and multimedia application grows, researchers are in pursuit of the next generation of satellite networks with low delivery time and high throughput. Therefore, modeling and analysis of delivery time, as a branch of satellite network research, are significant to the design and optimization of satellite networks for pervasive computing.

Because of the complexity of satellite networks, delivery time is affected by a variety of factors such as topology, routing, traffic distribution, etc. It is necessary to simplify the satellite network model to some extent. Chiajiu adopted M/M/1 Jackson queuing network to analyze the delivery time under the uniform distribution of traffic [2]. Based on this model, Jianhao introduced local distribution and decay distribution and compared the influence on the delivery time [3]. Fenge employed general stochastic Petri nets (GSPN) to model satellite networks and obtained delay-related performance curves using Petri net analysis software, which was then compared with results from simulations [4]. Yet, the Petri net state space would explode with the increase in complexity of topology and traffic load. In [5], delivery time under routing strategy based on capacity and flow assignment formula (CFAF) was studied. And the importance of queue delay in broadband LEO satellite networks was investigated in [6]. However, neither of them took traffic distribution and high traffic load into consideration.

Apparently, delivery time is directly affected by traffic source input and destination distribution. Due to the characteristics of satellite networks, traffic pattern likely exhibits spatiotemporal burst [7]. Most delivery time models, which assume Poisson input and uniform distribution, could scarcely describe the pattern.

Additionally, in a flow path, the output of upstream nodes becomes the input of downstream nodes, which forms correlation between them. Models based on the independence of individual queue fail to reproduce the correlation, undermining the accuracy of delivery time analysis.

Based on the above concern, a delivery time computational model for LEO satellite networks is proposed. In order to describe the spatiotemporal burst, Markov-modulated Poisson process (MMPP) [8] is taken as source input, while hot-spot distribution [9] is utilized for destinations. For any traffic flow between two stochastic nodes, a tandem queue model is established, of which the input is divided into a target flow and a cross-flow. By iteratively calculating the delay of target flow in each tandem queue, the end-to-end delivery time is obtained. Finally, a comparison between computational results and simulation results is conducted to validate the accuracy of the model.

In the next section, we establish the delivery time model for LEO satellite network. In Sect. 3, the algorithm of solving the model is developed. Performance evaluation is given in Sect. 4, while the conclusion is drawn in Sect. 5.

2 Model Establishment

2.1 The LEO Satellite Network

The LEO network consists of two parts, terrestrial segment formed by up and down user data links (UDL) and space segment formed by inter-satellite links (ISL). The packet delivery time is given as

$$T_{\text{packet}} = T_{\text{access}} + T_{\text{uplink}} + T_{\text{downlink}} + \sum_i^K T_{\text{sat_}i} + \sum_i^{K-1} T_{\text{cross_}i} \quad (1)$$

where T_{access} denotes the terminal access delay determined by the multi-access technique, T_{uplink} and T_{downlink} are the propagation delay of up and down links, T_{cross} is inter-satellite propagation delay, T_{sat} consists of process delay, queuing delay, and transmission delay, and K is the number of nodes in the path. Here, we mainly focus on the delay in space segment and make following assumptions:

1. The topology of the LEO satellite network is simplified as $N \times N$ torus. The propagation delays (T_{cross}) of ISLs are the same, and T_{sat} only consists of queuing delay and transmission delay.
2. Transmission paths are selected according to the principle of *column-first* routing: nodes always route packets along the column (intra-orbit) in which they are located toward the destination node D until they reach the D 's row (inter-orbit). Then, packets are sent along the D 's row until they reach D .

2.2 Source Traffic

The MMPP is capable of capturing both time-varying arrival rates and correlations between inter-arrival times and is easy to analyze and track so that it has been extensively used to model multimedia sources in broadband integrated services digital networks [10]. We take the dual-state MMPP as source traffic model, which can be denoted by the infinitesimal generator Q_j and the rate matrix Λ_j .

$$Q_j = \begin{bmatrix} -\sigma_{1j} & \sigma_{1j} \\ \sigma_{2j} & -\sigma_{2j} \end{bmatrix}, \Lambda_j = \begin{bmatrix} \lambda_{1j} & 0 \\ 0 & \lambda_{2j} \end{bmatrix} \quad (2)$$

where σ_{1j} and σ_{2j} are the transition rates of two states, λ_{1j} and λ_{2j} are, respectively, the Poisson arrival rates under two states. MMPP-2 has two important properties:

As r independent MMPP-2 traffic flows merge, the traffic descriptor (Q, Λ) of aggregated process can be written in terms of (Q_j, Λ_j) ($1 \leq j \leq r$) as

$$\begin{aligned} Q &= Q_1 \oplus Q_2 \oplus \cdots \oplus Q_r \\ \Lambda &= \Lambda_1 \oplus \Lambda_2 \oplus \cdots \oplus \Lambda_r \end{aligned} \quad (3)$$

where \oplus is Kronecker sum [8].

As a part of MMPP-2, (Q_j, Λ_j) traffic flow splits with probability p , and the traffic descriptor of splitting process can be expressed as $(Q_j, p \cdot \Lambda_j)$.

2.3 Destination Distribution

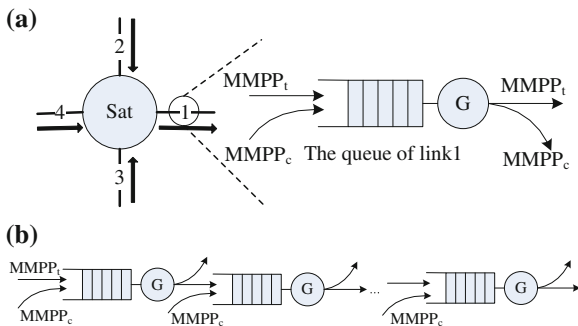
Since LEO satellite networks can provide global coverage for access service of pervasive computing, the traffic load of the network is influenced by the distribution of the terrestrial users, which usually forms hot spots in the network [7]. Besides, the *all-to-one* communication pattern, for example, the collection of all satellites' information from a ground station, will produce hot spots as well.

To reflect the effects from the distribution of the terrestrial users without introducing multi-access and handover problems, satellites are assumed to directly generate traffic to present the aggregation traffic from the ground. Meanwhile, hot-spot model is used to describe the spatial burst: set the location of the hot spot, to which the satellite nearest will be the hot spot. Nodes contact the hot spot with probability α and have a uniform communication with $1 - \alpha$.

2.4 System Model

Not only does a satellite generate traffic, but also they transfer flows from neighbors. Each node has four duplex links which, respectively, have one sending queue. If a packet arrives at a node, judge whether it has reached its destination. If

Fig. 1 System model:
a satellite link queue model;
b tandem queue model



so, remove the packet from the network otherwise add it into the corresponding queue.

As shown in Fig. 1a, we can divide the arrival traffic flow of the queue into two parts: the target flow $MMPP_t$, which is defined by the source and the destination, and the cross-flow $MMPP_c$, which aggregates other flows except the target flow and impacts the queue delay of the target flow. Assume that the service times of the queues are independent and identically distributed (i.i.d.) with distribution function $\tilde{H}(x)$ whose Laplace–Stieltjes transform (LST) and mean service time are, respectively, $H(s)$ and h . The capacity of the queue buffer is infinite so that no overflow exists. Packets are served with *first-come-first-serve* (FCFS) strategy.

According to the *column-first* routing, the path of the target flow can be decided, which consists of $K(1 \leq K \leq N)$ nodes forming tandem queues. In Fig. 1b, besides the target flow, each queue has cross-flows to arrive and departure. Therefore, the end-to-end delivery time computation is transformed into the delay that flow $MMPP_t$ experiences in the tandem queues. First, we need to decide the descriptors of $MMPP_t$ and $MMPP_c$ of each link. Then, we can calculate the queuing delay of the target flow in each queue, from which the end-to-end delay can be obtained.

3 Model Solving

3.1 Traffic Load Distribution

In order to get the parameters of cross-flows of necessary links, we first obtain the descriptors under uniform pattern and hot-spot pattern, respectively, and then aggregate the descriptors for the delivery time computation.

Under uniform pattern.

Lemma 1 *Node i is a stochastic node in the $N \times N$ network. Let $n_i(k)$ be the number of nodes that are k hops away from node i [11].*

If N is odd,

$$n_i(k) = \begin{cases} 1, & k = 0 \\ 4k, & 0 < k \leq (N-1)/2 \\ 4(N-k), & (N-1)/2 < k \leq N-1 \end{cases} \quad (4)$$

If N is even,

$$n_i(k) = \begin{cases} 1, & k = 0 \\ 4k, & 0 < k \leq N/2 \\ 4k-2, & k = N/2 \\ 4(N-k), & N/2 < k < N \\ 1, & k = N \end{cases} \quad (5)$$

Lemma 2 *In the $N \times N (N \geq 3)$ torus network, where each node generates traffic at the rate R , under uniform communication pattern with the column-first routing, the average link load is presented as: if N is odd, the average loads of links are the same as $RN/8$; if N is even, the average loads of two column (row) are different, given as $R(N \pm 2)/(8 - 8/N^2)$.*

Using Lemma 1, the proof of Lemma 2 is straight, and it is skipped due to the constraint of the length. Based on Lemma 2, if the traffic generation of a node follows MMPP_s with the descriptor (Q_s, Λ_s) or $(\sigma_1, \sigma_2, \lambda_1, \lambda_2)$, $R = (\lambda_1\sigma_2 + \lambda_2\sigma_1)/(\lambda_1 + \lambda_2)$, according to the properties of MMPP , assume that the arrival process of the link is MMPP_u with (Q_u, Λ_u) , which is aggregated by $\gamma (\gamma \in R^+)$ MMPP_s flows. Then, we have Theorem 1.

Theorem 1 *If N is odd, $\gamma = N/8$; if N is even $\gamma = (N \pm 2)/(8 - 8/N^2)$. If $\gamma \leq 1$, $Q_u = Q_s$, $\Lambda_u = \gamma\Lambda_s$; If $\gamma > 1$ and $\gamma = n + \beta (n \in \mathbb{N}, 0 \leq \beta < 1)$, $Q_u = \underbrace{Q_s \oplus \cdots \oplus Q_s}_{n+1}$, $\Lambda_u = \underbrace{\Lambda_s \oplus \cdots \oplus \Lambda_s}_n \oplus (\beta\Lambda_s)$.*

Proof According to Lemma 2, the relation between the link arrival process and the source process is defined. If $\gamma \leq 1$, from MMPP 's second property, $Q_u = Q_s$, $\Lambda_u = \gamma\Lambda_s$; otherwise MMPP_u can be treated as the aggregation of n MMPP_s s and one partial MMPP_s . Then, we get the upper results using Kronecker sum, of which the computation can be simplified with the dual-state approximation [12].

Under hot-spot pattern. Under the hot-spot communication pattern, with *column-first* routing, flows gather in the columns and the hot spot's row. Let N be odd. Then we can build the coordinate taking the hot spot as the origin with nodes' interval being one unit. As $x_i, y_i \in [0, (N-1)/2]$, the load of the link $(x_1, y_1) \rightarrow (x_2, y_2)$ has two cases: if $y_1 = y_2 + 1$ and $x_1 = x_2$, the hot-spot load MMPP_h can be seen as the aggregation of $(N-1)/2 - y_1 + 1$ MMPP_s s; if $y_1 = y_2 = 0$ and $x_1 = x_2 + 1$, assume MMPP_v is the aggregation of vertical link load, the MMPP_h consists of $N/2 - x_1 + 1$ MMPP_v s. The cases in other quadrants are the same. All these computations can be easily completed with the approximation.

Cross-traffic load. Consider the combination of the uniform pattern and the hot-spot pattern. First, we have the stability condition of the network based on above:

Theorem 2 *If the probability of the hot-spot pattern is α , that of uniform pattern is $1 - \alpha$ and the link capacity is C , the stability condition of the network is: if N is odd, $R < \frac{8C}{4N^2\alpha + (1-5\alpha)N}$; if N is even, $R < \frac{8C(N^2-1)}{(N+2)(4N^2\alpha-5\alpha+1)N^2}$.*

The cross-flow of each link consists of uniform flows or hot-spot flows. If the destination is the hot spot, $MMPP_c$ is composed of other hot-spot flows and uniform flows. Otherwise, the $MMPP_c$ of the links in the hot-spot paths consists of hot-spot flows and uniform flows while that of other links only consists of $MMPP_u$. In this way, we can decide the cross-flow of each link in the transmission path.

3.2 Single Queue Solving

The single queue in the network can be modeled as $MMPP_t + MMPP_c/G/1$. In this part, the computational methods of departure interval moments of the aggregated flows and the delay of target flow in a single queue are given, on which the method of calculating departure interval moments of the target flow based is presented in the next part.

Departure interval moments of the aggregated flows. Assume $\{\tau_n : n \geq 0\}$ is the successive epochs of departure with $\tau_0 = 0$. Let L_n be the number of packets in the queue right after τ_n and A_n be the number of arrival packets between the $(n - 1)$ th and the n th departure. Then, $L_n = \max(L_{n-1}, 0) + A_n$. Define J_n is the corresponding state of the aggregated flows $MMPP-m$. Then, the triple $\{(L_n, J_n, \tau_{n+1} - \tau_n) : n \geq 0\}$ forms the semi-Markov sequence with the transition matrix:

$$\tilde{Q}(x) = \begin{bmatrix} \tilde{B}_0(x) & \tilde{B}_1(x) & \tilde{B}_2(x) & \cdots \\ \tilde{A}_0(x) & \tilde{A}_1(x) & \tilde{A}_2(x) & \cdots \\ 0 & \tilde{A}_0(x) & \tilde{A}_1(x) & \cdots \\ \cdots & \cdots & \cdots & \ddots \end{bmatrix} \tag{6}$$

where $\tilde{A}_n(x)$ and $\tilde{B}_n(x)$ are the $m \times m$ matrixes of mass functions. Assume $A_n(x)$ and $B_n(x)$ are the LST of $\tilde{A}_n(x)$ and $\tilde{B}_n(x)$, respectively. Let $\mathbf{x} = (\mathbf{x}_0, \dots, \mathbf{x}_k, \dots)$ denote the vector of the stationary probability of $\tilde{Q}(\infty)$, in which the j th element of \mathbf{x}_k $x_{k,j} = \lim_{n \rightarrow \infty} P\{L_n = k, J_n = j\}$. And \mathbf{x}_k satisfies

$$\mathbf{x}_k = \mathbf{x}B_k(0) + \sum_{v=1}^{k+1} \mathbf{x}_v A_{k+1-v}(0) \tag{7}$$

\mathbf{x}_0 can be calculated referring to [8] so that the queue size distribution after a departure can be derived.

Let T_i be the time between the i th and the $i + 1$ th departure and $\tilde{D}^n(x_1, \dots, x_n)$ denote the probability generating function of the joint distribution of n successive intervals T_i with LST $D^n(s_1, \dots, s_n)$. Then,

$$D^1(s) = H(s) - sH(s)\mathbf{x}_0(sI - D_0)\mathbf{e} \quad (8)$$

where $D_0 = Q - \Lambda$, \mathbf{I} is a $m \times m$ unit matrix and \mathbf{e} is $m \times 1$ column vector of all ones. The n th moment of T_i is

$$E[T_i^n] = (-1)^n \frac{\partial^n}{\partial s^n} D^1(s)|_{s=0} \quad (9)$$

and the covariance of two successive intervals is

$$\text{Cov}(T_i, T_{i+1}) = \frac{\partial^2}{\partial s_i \partial s_{i+1}} D^2(s_i, s_{i+1})|_{s_i=s_{i+1}=0} \quad (10)$$

Delay of the target flow. Let $\tilde{W}(x) = \{W_1(x), \dots, W_m(x)\}$, where $W_j(x)$ is the joint probability of the arbitrary time when the arrival state is j and the waiting time of the arrival packet at the moment is no more than x . The LST of $\tilde{W}(x)$ is $W(s)$, which can be calculated as [8]:

$$W(s) = s(1 - \rho)\mathbf{g}[sI + Q - \Lambda(1 - H(s))]^{-1} \quad (11)$$

where $s > 0$, $\{\mathbf{g}\}$ is the vector of stationary probability of the transition matrix G in the busy period, and $\rho = \lambda h$. The n th moment of the packet waiting time vector \mathbf{w} :

$$E[\mathbf{w}^n] = (-1)^n \frac{\partial^n}{\partial s^n} W(s)|_{s=0} \quad (12)$$

Then the average waiting time of the target flow is

$$w_i = \frac{1}{\lambda_i} E[\mathbf{w}] \Lambda(i) \mathbf{e} \quad (13)$$

. Thus, the average delay in a single queue is $T_{\text{sat}} = w_i + h$.

3.3 Tandem Queue Solving

Departure interval moments of the target flow. In order to get the input parameters of the target flow for the next queue, the separation of target flow from the cross-flow is necessary. We replace the queue model $\text{MMPP}_i + \text{MMPP}_c/G/1$ with $\text{MMPP}_i/G/1$ by changing the service time to approximate the influence from the cross-flow. Since packets of the target flow is able to be served directly only if no packets of the cross-flow are in the front, the target flow's probability of being

Table 1 The algorithm of the delivery time

Algorithm 1 Tandem_Delay()	
1	Determine the K links of the target flow path
2	for $i = 1 : K$
3	Compute the link descriptor (Q_c, Λ_c) ;
4	Compute the aggregated flow's descriptor (Q, Λ) ;
5	Compute the delay T_{sat} ;
6	Compute the moments of the inter-departure time;
7	Compute the input parameters of the next queue;
8	$T_{\text{tandem}} = T_{\text{tandem}} + T_{\text{sat}}$;
9	end

served without waiting is $p = 1 - \rho_c$ where $\rho_c = \lambda_c h$. Thus, the effective service time for the target flow is

$$h_t = ph + \sum_{i=1}^{\infty} i(1-p)^i h = \frac{h}{1-\rho_c} \quad (14)$$

Then, using the method in Sect. 3.2, $E[T_i]$, $E[T_i^2]$, $E[T_i^3]$ and $\text{Cov}(T_i, T_{i+1})$ of the queue MMPP _{i} /G _{i} /1 can be obtained, which are necessary for the moment matching. The descriptor $(\sigma_1, \sigma_2, \lambda_1, \lambda_2)$ of the target flow in the next queue is derived from these inter-departure moments through moment matching, which is described in detail in [13].

Delivery time of the target flow. According to the assumption, the delivery time of the target flow consists of the delay packets experience in the tandem queues. We need to iteratively sum up the average queuing delays and transmission delays. The complete steps of algorithm are given in Table 1.

After T_{tandem} is calculated, T_{packet} can be obtained according to (1) and the assumptions.

4 Performance Evaluation

To validate the accuracy of delivery time computational model, the comparison between results of simulation in ns2 and that of computation in MATLAB is conducted. In order to get the reliable statistics, the simulation is repeated independently for 50 times with 96 % confidence interval.

The simulation parameters are set as following: the capacity of ISLs $C = 4$ Mbps; the link propagation delay $T_{\text{cross}} = 10$ ms; the link error rate is zero; the buffer of the link queue is infinite so that there is no packet loss during simulation; the source input for each node is MMPP-2 $(\sigma_1, \sigma_2, \lambda_1, \lambda_2)$ where $\sigma_1 = \sigma_2 = 0.01$, $\lambda_1 = 2\lambda_2$, $\rho = R/C$, $\rho \in [0.1, 0.9]$ and the lingering time of each state follows exponential distribution with the mean $1/(1 - 0.01) = 1.0101$ s; the packet length follows three different distributions: deterministic, exponential, and k -stage Erlang, with the average value 1 KB ($h = 2.048$ ms).

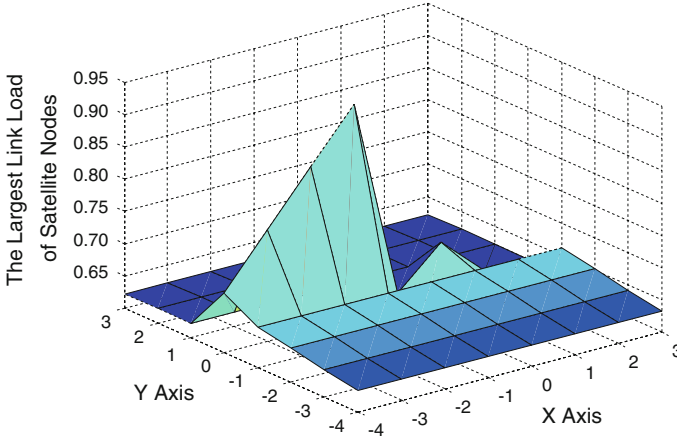


Fig. 2 The distribution of the largest link loads as $\rho = 0.5$. The (x, y) is the logical coordinate of satellite node taking the hot-spot node as $(0, 0)$. The z -axis denotes the largest link load of each satellite node

The satellite network, similar to Iridium, is composed of an 8×8 torus with 780 km height, whose period is about 100 min set as the simulation duration. The satellite nearest to the location $(45^\circ\text{N}, 25^\circ\text{E})$ will become the hot spot with $\alpha \in [0.01, 0.09]$. For the controllable sending rates, the transport layer has no acknowledges and flow control; packets are deleted after arrival at their destinations.

First, we examine the results of traffic load distribution with $\alpha = 0.02$. Using Theorem 2, the stability condition is $\rho < 0.65$. In Fig. 2, as $\rho = 0.5$, the largest link load of the four links of each node is presented with the coordinate of the hot spot $O(0, 0)$. For N is even, the two sides of the hot spot are not symmetric, leading to the unbalanced load distribution. The node $(-1, 0)$ has the bottleneck link with load near 0.95.

Then, the descriptor of the cross-flow of each link in the path decided by the source and the destination can be given. For example, in the path from $A(-4, -4)$ to O , there are eight cross-flow descriptors according to the routing. Using these parameters, we get the average queuing delay through Algorithm 1, as depicted in Fig. 3. The first six computational results are right in the confidential interval of the simulation results, while the last two values are a little higher. This is because, under a high cross-flow load, the inter-departure interval computing result of the target flow is lower than that of the simulation, leading to the deviation of the queue delays. Additionally, the cumulative errors of the descriptors are also responsible for the variation.

In Fig. 4, under varied source traffic loads, the delivery time from A to O is illustrated by contrast with that from B to O with six hops. Due to the extra hops and higher cross-flow load, the delivery time of A is longer and grows faster than that of B . As $\rho = 0.4$, the value of A starts to deviate from the simulation result;

Fig. 3 The average queuing delays of the links in the path. The x -axis denotes the number of each queue in the path, and the y -axis is the average queuing delay of each corresponding queue

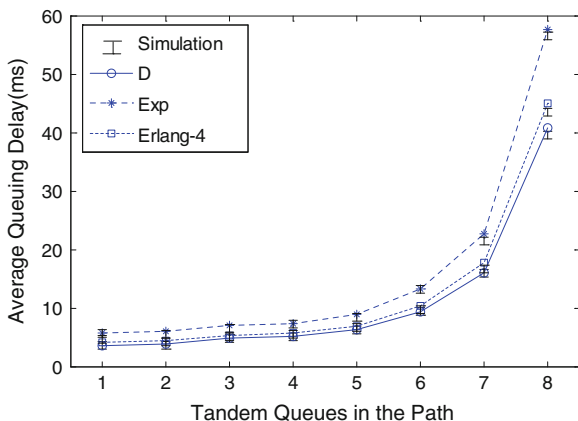
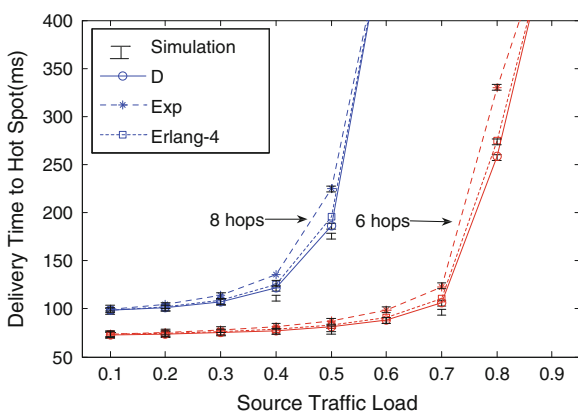


Fig. 4 The delivery times under varied loads. The x -axis denotes the varied source traffic load, and the y -axis is the delivery time from source to the hot spot. The blue lines refer to the 8-hop path ($A-O$), and the red lines are the 6-hop path ($B-O$)

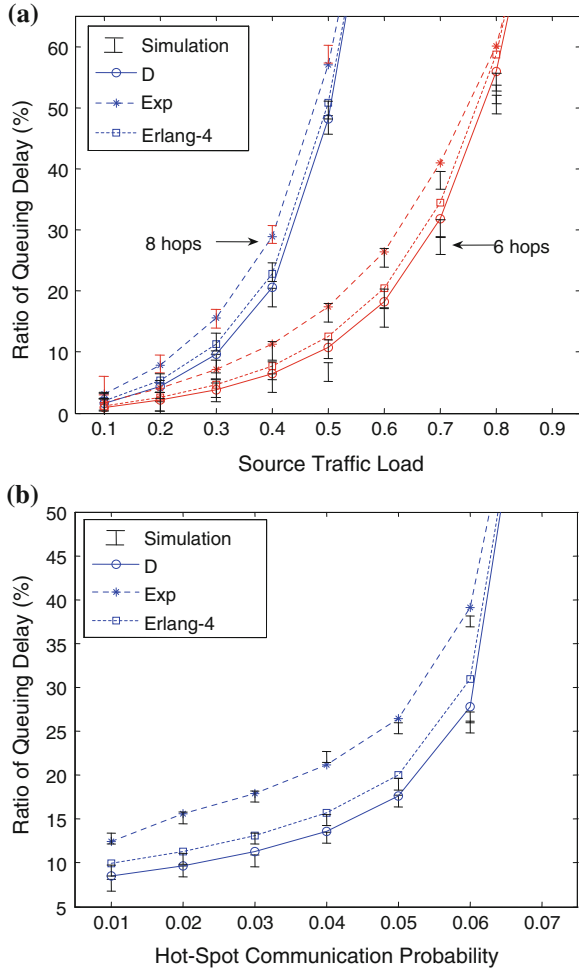


after $\rho \geq 0.6$, the queue in the path becomes unstable and the delays soar unboundedly, which are actually invalid values in the computation. But this situation is not for B until $\rho \geq 0.7$.

In order to demonstrate the influence of traffic load on the delivery time more clearly, the ratios of the queue delay over the delivery time of A and B are shown in Fig. 5a. In contrast to the conclusion in [6] that the queue delay can be ignored in the satellite network with high-capacity links, the ratios keep over 10 % under most traffic loads. In addition, when we raise the probability of hot-spot communication, the ratio of A increases dramatically as depicted in Fig. 5b. It can be inferred that though the high-capacity link reduces the transmission delay, it fails to compensate the boost of queuing delay under a certain traffic distribution.

As we can see, the computational results are consistent with the simulation both in the trend and in an acceptable margin of deviation. Moreover, the computational time is far below that consumed by simulation. Therefore, for a large number of discrete events, it is reasonable to replace simulations with this model. However,

Fig. 5 a The queuing delay ratio under varied source traffic load. The x -axis donates the varied source traffic load, and the y -axis is the ratio of queuing delay over delivery time from source to the hot spot. The blue lines refer to the 8-hop path ($A-O$), and the red lines are the 6-hop path ($B-O$). **b** The queuing delay ratio under varied probability of nodes communicating with the hot spot. The x -axis donates the varied hot-spot communication probability, and the y -axis is the ratio of queuing delay over delivery time from A to O .



there are still a lot of necessary improvements for this model, for example, the accuracy guarantee under high traffic load, the introduction of asymmetric topology, and the metrics such as throughput and packet loss rate. We will accomplish these features in our future works.

5 Conclusion

LEO satellite networks, with the ability to provide global broadband access service, are considered as the complement and extension of the terrestrial networks when developing the ubiquitous infrastructures of pervasive computing. Among the requirements of pervasive computing, the delivery time is essential for users to

access computing resource. To investigate the time constraint, a delivery time computational model for LEO satellite networks is proposed. With source inputs of MMPP and a hot-spot distribution of destinations, a tandem queue for the target flow is established with cross-flows. The descriptors of cross-flows are calculated through the aggregation of traffic flows, according to the topology and the routing. For the target flow, the departure interval moments are calculated and fitted for the next link as input parameters, from which the queuing delay under cross-flows is iteratively obtained. The comparison between computational results and simulation results demonstrates that this model makes a good depiction of the influence by the traffic pattern with satisfying accuracy, which would be valuable for the research of routing strategies and congestion control in satellite networks.

References

1. Taleb T, Kato N, Nemoto Y (2005) Recent trends in IP/NGEO satellite communication systems: transport, routing, and mobility management. *IEEE Wirel Commun Mag* 12(5):63–69
2. Wang CJ (1995) Delivery time analysis of a low earth orbit satellite network for seamless PCS. *IEEE J Sel Areas Commun* 13(2):389–396
3. Hu JH, Wu SQ, Li LM (1999) Performance analysis for LEO/MEO mobile satellite communication systems with intersatellite link. *ACTA Electronica Sinica* 27(11):65–68
4. Wu FG, Sun FC, Yu K, Zheng CW (2005) Performance evaluation on a double-layered satellite network. *Int J Satell Commun Network* 23(6):359–371
5. Yeo BS (2002) An average end-to-end packet delay analysis for LEO satellite networks. *IEEE vehicular technology conference*, pp 2012–2016
6. Jurski J, Wozniak J (2009) Routing decisions independent of queuing delays in broadband LEO networks. *IEEE global telecommunications conference*
7. Mohorcic M, Svigelj A, Kandus G, Hu YF, Sheriff RE (2003) Demographically weighted traffic flow models for adaptive routing in packet-switched non-geostationary satellite meshed networks. *Comput Netw* 43:113–131
8. Fischer W (1992) The markov-modulated poisson process (MMPP) cookbook. *Perform Eval* 18:149–171
9. Min GY, Wu YL, Mohamed OK, Yin H, Li KQ (2009) Performance modeling and analysis of interconnection networks with spatio-temporal bursty traffic. *IEEE global telecommunications conference*
10. Ferng HW, Chao CC, Peng CC (2007) Path-wise performance in a tree-type network: per-stream loss probability, delay, and delay variance analyses. *Perform Eval* 64(1):55–75
11. Sun J, Modiano E (2004) Routing strategies for maximizing throughput in LEO satellite networks. *IEEE J Sel Areas Commun* 22(2):273–285
12. Heindl A (2003) Decomposition of general queuing networks with MMPP inputs and customer losses. *Perform Eval* 51:117–136
13. Ferng HW, Chang JF (2001) Connection-wise end-to-end performance analysis of queuing networks with MMPP inputs. *Perform Eval* 43:39–62

How People Use Visual Landmarks for Locomotion Distance Estimation: The Case Study of Eye Tracking

Huiting Zhang and Kan Zhang

Abstract Research has been focusing on how people navigate in the virtual space since the technology of virtual reality was developed. However, not enough has been known about the process of the virtual space cognition. During locomotion, distance could be visually accessed by integrating motion cues, such as optic flow, or by the self-displacement process in which people compare the change of their self-position relative to individual identifiable objects (i.e. landmarks) in the environment along the movement. In this study, we attempted to demonstrate the effect of the later mechanism by separating the static visual scenes from the motion cues in a simulated self-movement using a static-frame paradigm. In addition, we compared the eye tracking pattern in the static scene condition (without motion cues) with the eye tracking pattern in the full visual cue condition (with motion cues). The results suggested that when only static visual scenes were available during the simulated self-movement, people were able to reproduce the traveled distance. The eye tracking results also revealed there were two different perceptual processes for locomotion distance estimation and it was suggested that locomotion distance could be estimated not only by optic flow as we already knew, but also by the self-displacement process from the visual static scenes.

Keywords Landmarks · Locomotion distance estimation · Eye tracking

H. Zhang (✉) · K. Zhang
Institute of Psychology, Chinese Academy of Sciences, Beijing, China
e-mail: zhanghuiting@psych.ac.cn

K. Zhang
e-mail: zhangk@psych.ac.cn

H. Zhang
University of Chinese Academy of Sciences, Beijing, China

1 Introduction

Visual perception has long been a critical subject in the spatial cognition and virtual reality research. Estimation of the locomotion distance is important for navigation and spatial learning. When people are moving in the real world, distance estimation process happens naturally throughout the trip that the information from vision [1], proprioception, motor commands [2, 3] and vestibular sense [4] is received and integrated together. However, in the virtual reality, not only is the visual information different from the real world, but the body sense is also largely reduced. In previous studies, humans were found to be able to judge the distance in a virtual environment or estimate the traveled distance regardless of a certain amount of bias via visual information as the main perceptual cues. In this study, we further divided the visual information during a simulated movement into motion cues and static scenes, and attempted to prove the existence of two different perceptual processes based on these two kinds of visual information.

2 Related Works

Humans were found to be able to estimate their movements and traveled distance depending only on the visual information [5–7]. In a visually full-cue environment, there are two basic mechanisms that visual information can be used to localize oneself in an environment. The estimation of locomotion distance could be done by integrating the visual motion cues about the direction and speed of one's movements over time [8, 9]. For example, in order to judge the distance traveled on the basis of optic flow, the observer can first estimate the velocity and duration of the self-motion [1, 10]. And then the velocity could be integrated over time to determine the distance that the observer has traveled. Estimation of distance traveled based on optic flow, such as in a texture environment, has been demonstrated with various tasks, such as distance discrimination [11], distance adjustment [12, 13] and distance reproduction [14]. In all cases a linear relationship between the perceived and the actual distance was observed, but with a consistent undershooting of absolute magnitude.

Vision can also be used to directly determine one's position relative to individual identifiable objects in the environment [15] based on a place/scene recognition process. For example, landmarks, defined as visible, audible, or otherwise perceivable objects which are distinct, stationary, and salient [14], can be treated as reference about one's position within the perceptual range, and can support self-localization process. It is known that people have the ability to judge the egocentric distance between oneself and an object in different static environments [16, 17]. Much research has been done to examine how humans perceive egocentric distance when viewing a target from a fixed viewpoint [18]. Perceived distance, for the most part, is also linearly correlated to physical distance though

consistently underestimated as the simulated distances increased [13, 19]. Therefore, traveled distance can also be assessed by subtracting the egocentric distance from the observer to a certain perceivable landmark between the static scenes along the trip, between the end of the trip and the start of the trip for instance.

Very few studies to our knowledge investigated the effect of only static visual cues on the estimation of the distance traveled during locomotion. Lappe et al. [15] studied distance judgment involving static scenes using a virtual environment of a hallway with randomly colored panels on both walls. Participants were first visually moved a certain distance and then asked to adjust a target in a static scene for the same distance or first shown a target in a static scene and then asked to translate the same distance with active or passive simulated movement. However, in this experiment, the static visual scenes were only used for estimating the egocentric distance instead of the locomotion distance. Thus it did not address the question on whether locomotion distance estimation is possible with static visual information alone.

Besides, studies on eye movements during locomotion understand well about how the visuomotor and vestibule-motor systems function and interact [20]. However, there was little research focusing on the eye tracking with the visual cues, specifically on landmarks, in the field of locomotion distance estimation either in the real world or in virtual realities, partly because the technological constrains and the challenges for recording and the data analysis of the eye movement on dynamic stimuli in 3D space.

In this study, we sought to compare the effect of static visual information (without motion cues) on locomotion distance estimation with the full cue visual condition (with motion cues), in other word, to investigate of the effectiveness of static-scene mechanism. The virtual environment built in this experiment was a tunnel with several distinctive, identifiable landmarks on the walls, floor and ceiling. The perceptual landmarks can first provide unique patterns of optic flow during locomotion and thus can be used for distance estimation using the motion-based mechanism. At the same time, distinctive perceptual landmarks can also function as self-localization cues to determine one's position at a given moment, and therefore supporting a mechanism based on the static scene. To separate the static scenes from the motion cues, a static-frame paradigm was first developed and used. The detailed description would be given in the method part. Furthermore, to validate whether people actually use different mechanisms when different cues were available, we also recorded their eye movement throughout the experiment. In our hypothesis, when full visual information is available during the locomotion, people follow the position of the perceptual landmarks throughout the trip to access the optic flow and integrate the traveled distance and their gaze points should be moving smoothly and highly close to the position of the landmarks. However, when the motion cues are eliminated and only discrete static scenes are provided, their gaze points should be moving inconsecutively or jumpily according to the discontinuously position change of the landmark between the static scenes.

3 Method

3.1 Subjects

Sixteen undergraduate and graduate students (9 males, and 7 females) participated in this experiment. All participants had normal or corrected-to-normal vision and signed the consent form and were paid for their participation.

3.2 Apparatus and Virtual Environment

The experiment was conducted on a PC, running a C++ program using open GL. Participants were seated in a dimly illuminated room 60 cm from a 17-in display monitor, rendered at 72 Hz refresh rate, a resolution of 1,024 × 768 pixels, and a graphical field of view of 40° × 30°. The eye movement was recorded by Tobii T120 Eye Tracker whose sampling rate is 120 Hz.

The virtual environment was a hallway-like tunnel, 2 m wide and 3.2 m high (the simulated eye height was 1.6 m). The floor, ceiling and walls of the tunnels were set with different visual cues and there were different visual environments for learning phase and distance reproduction phase in each trail. In the learning phase (see Fig. 1a), all four walls were the same solid gray color, with four objects in different shapes, one on each wall, in the order of a blue rectangle on the left wall, a red circle on the floor, a green triangle on the right wall, and a yellow star on the ceiling. The positions of the four shapes were fixed for all trials. They were placed in the middle of the wall about 0.625, 6.25, 11.875 and 17.5 m from the starting point, respectively. While in the reproduction phase, another environment was used to prevent people from using simple scene-match strategy. Four walls of the new tunnel were painted with a green and dense-patterned texture. In addition, 12 yellow diamond shapes were used as perceptual landmarks in this reproduction environment, with fixed locations different from each of the four objects used in the learning phase, three on each wall (see Fig. 1b). The order and position of these shapes were randomly chosen and kept the same for all trials.

3.3 Design and Procedure

In each trial, participants were asked to first watch a simulated self-movement along the center line of the tunnel for a certain distance (the learning phase) and then to reproduce the distance with another simulated movement in a different speed (the reproduction phase). Each simulated self-motion was initiated by pressing the “SPACE” key on the keyboard by the participants, followed by a fixation screen for 500 ms before the test stimuli appeared. In the learning phase,

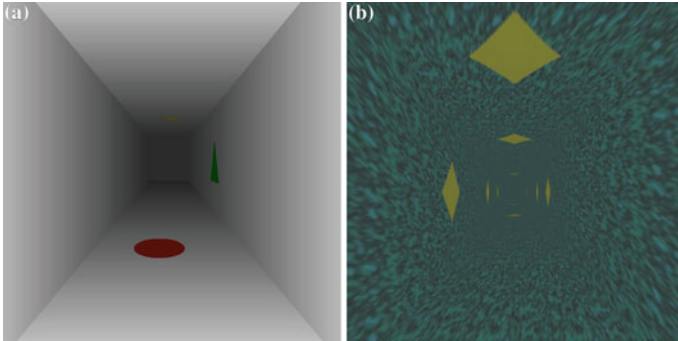


Fig. 1 The illustration of the tunnel in **a** the learning phase and in **b** the reproduction phase

the movement stopped automatically. In the reproduction phase, the participants needed to decide when/where to stop the movement by pressing the “SPACE” key again. Participants were told the locomotion speed was randomly chosen and were different in the learning and reproduction phase to prevent them from counting the time. However, for the analysis of the eye tracking data, actually we kept the speed constant as 1.25 m/s in the learning phase, and 1.5 m/s in the reproduction phase. The corresponding learning duration was 4.8–9.6 s, and the response duration was determined by the participants.

A static-frame paradigm was used to manipulate the availability of the motion cues’ during the movement. In the learning phase, while the simulated movement was continuous, only a sequence of static scenes at given instances were provided to the participants, mimicking a walk in the dark tunnel where a strobe light periodically illuminated the surroundings. Every static frame was presented for 100 ms, and participants could see where they were in the tunnel at that moment. The interval between every static scene was completely dark, and the durations lasted for 0 or 1,000 ms. When the interval is 0, though the physical stimuli were a series of static frames, the apparent motion existed and continuous visual scenes would be perceived and the 0 ms interval condition served as a full-cue condition (with both motion cue and static scenes). When the interval was 1,000 ms, the blank was long enough for participants to notice, and the discrete static scenes would be perceived. During the reproduction phase, both the movement and the visual scene were continuous (like the tunnel was continuously illuminated).

A two-factor within-subject design was used, with 3 (distance: 6, 9 or 12 m) \times 2 (interval durations: 0 or 1,000 ms) and each condition was tested for three times in three blocks. All trials were randomized within a block. Participants were not given any feedback about their performance. The whole experiment lasted about 20 min.

4 Results

Using dynamic stimuli in the eye movement study was challenging from the technical aspect for the eye movement would be more complicated than the eye movement on an image or a sentence. It was reasonable that the eye tracker could not record as complete data as it does when it is used for pictorial material or web pages. In this case, we used data that had more than 50 % sampling results (data of one participant was eliminated from the following analysis). Since the original sampling rate is 120 Hz, we believed the 50 % sampling was acceptable.

Participants' eye movements were recorded for the whole experiment. In this study, since the stimuli were changing continuously and rapidly, we currently focused on the gaze data in the learning phase, to investigate which position on the screen they looked at to perceive the distance. The distance reproduction performance and eye tracking data were analyzed and reported separately.

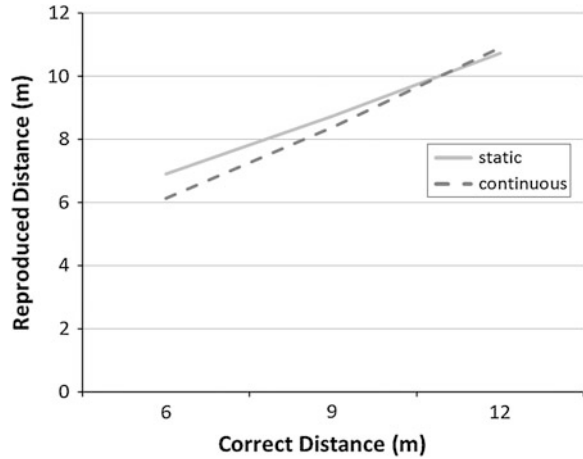
4.1 Distance Reproduction

One data point was excluded from the analysis whose reproduction distance was 0.225 m. For the shortest distance to be estimate was 6 m, we believed this extreme data was from accidentally pressing the SPACE key too quickly in the reproduction phase.

A repeated measure ANOVA (3 correct distances \times 2 intervals) was conducted on the reproduced distance. To better understand the results, we referred 0 ms interval condition as the continuous condition and the 1,000 ms interval condition as the static scene condition. Only the main effect of the correct distance was significant ($F_{(2,28)} = 104.338$, $p < 0.001$, see Fig. 2). Participants tended to overestimate the distance of 6 m ($M = 6.303$, $SE = 0.352$) and underestimate the distance of 9 m ($M = 8.355$, $SE = 0.428$) and 12 m ($M = 10.537$, $SE = 0.591$). Neither the effect of the interval ($F_{(1,14)} = 0.912$, $p = 0.356$) nor the interaction was significant ($F_{(2,28)} = 0.120$, $p = 0.888$).

In addition, if the participant's estimation is accurate, the reproduction distance should be highly correlated with the correct distance, and the slope of the reproduction distance in the function of the correct distance should be close to 1. Therefore, we ran linear regressions for the reproduction distance in relationship with the correct distance for each participant. And the t test on the slope between two interval conditions ($M_{\text{static}} = 0.692$, $SE = 0.058$; $M_{\text{continuous}} = 0.719$, $SE = 0.093$) failed to reveal any significant differences ($t_{(14)} = -0.294$, $p = 0.774$).

Fig. 2 Reproduction distance as the function of the correct distance in different interval conditions

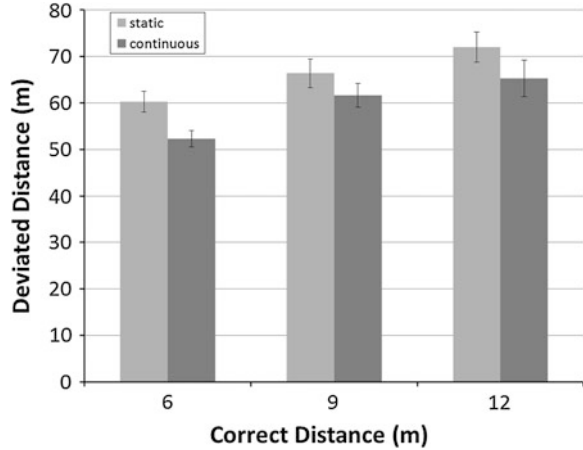


4.2 Gaze

The raw gaze data was recorded in form of the x-y coordinates of the screen. In order to decide whether the participants were looking at the specific landmark, we need to compare the gaze points of the participants with the location of the landmarks along the simulated movements. First, the simulated self-movements in the interval-0 condition were divided into several 100 ms-long segments, and the first frame of these 100 ms-long segments were used as the key frames. Next, the coordinates of the center of the landmarks in the key frames were recorded and set as the target points. Then the distance (referred to “the deviation distance” hereafter) from participants’ gaze point to the simultaneous target point was calculated as the dependent variable. The main consideration of this transformation was that both the location and the size of the landmarks were changing continuously and it would be a time-consuming process to circle the area-of-interest frame by frame for each participant. Compared to the effort, the benefit from this finest coding would be trivial since the pixel change of the location and size of the landmarks between two sequent frames was really small. Therefore, we used the deviation distance for the following analysis.

We ran a repeated measure ANOVA (3 correct distances × 2 intervals) on the deviation distance. First, as the same as the result of the reproduced distance, the main effect of distance was significant ($F_{(2,28)} = 18.537, p < 0.001$, see Fig. 3). It was suggested that participants better followed the location of the landmarks when the distance was short ($M_{-6} = 56.290, SE = 1.547; M_{-9} = 65.006, SE = 2.409; M_{-12} = 68.656, SE = 1.962$). Besides, the effect of the interval was also significant ($F_{(1,14)} = 4.566, p = 0.051$) that when the interval was 0 ($M = 66.239, SE = 2.395$), participants were more likely to follow the location of the landmark than they did when the interval was 1,000 ms ($M = 59.729, SE = 2.032$). No interaction effect of these two factor was significant ($F_{(2,28)} = 0.207, p = 0.814$). In addition, we did the same analysis on the standard deviation of the deviation distance

Fig. 3 Deviation from the center of the landmark as the function of the correct distance in different interval conditions



to investigate the constancy of the eye following behavior on the landmarks in different interval conditions. However, the only significant result found was the effect of the correct distance ($F_{\text{distance}(2,28)} = 3.962, p < 0.05$; $F_{\text{interval}(1,14)} = 0.105, p = 0.750$; $F_{\text{interaction}(2,28)} = 2.146, p = 0.136$).

5 Discussion

Two questions were addressed in this experiment: whether people were able to reproduce the distance during simulated locomotion when the motion cue was removed, as in the interval-1,000 condition; and how they used the perceptual landmarks when they perceived the distance.

In a distance reproduction task, participants first visually traveled a given distance in an environment with only fixed landmarks, and then they were asked to reproduce this distance in a different environment. Two alternative strategies were eliminated or restricted in our experiment. People could not reproduce the distance by just visually matching the visual scene at the end of each movement from the learning environment to the reproduction environment. And they could not reproduce the distance by just counting the time they spent in the learning phase and reproducing the duration in the reproduction phase either, otherwise, their reproduced distance would be shorter than the correct distance in each distance condition because the reproduction speed was always faster than the learning speed.

There was a strong distance effect, that people slightly overestimated the shorter distance and increasingly underestimated the longer distances. This result was consistent with previous finding in the distance estimation studies with blind walking tasks and distance estimation tasks in the virtual reality [11, 18, 21]. The range of the underestimation and overestimation was also comparable to the results from these previous studies.

However, the results from the reproduction distance revealed no difference under the different interval conditions. As we expected, when the interval was 0 ms, participants reported to sense a smooth locomotion in debriefing, while when the interval was 1,000 ms, they could clearly tell there were blanks from time to time and perceived a discontinuous, jump-like movement. The change of the environments and the moving speed from learning to testing required the participants to actually estimate the distance without using other strategies. Therefore, it was suggested that participants could estimate the locomotion distance when they motion cues were reduced, and their estimation was as good as the estimation in the full visual cue condition.

On the basis of the reproduction performance, we further explored the gaze data of the participants when they perceived the distance in the learning phase. The hypothesis was, when the interval was 1,000 ms and there was an apparent black between static frames/scenes, if people still can use the motion-based mechanism, for example to estimate the velocity, their gaze point should be closely and smoothly follow the position of the landmarks, that would be the actual position of the landmarks when the static frame was presented or the imaginary position of the landmarks during the blank. If they use the static mechanism, they should look at the position of the landmarks only when the static scene is presented and might use the saccadic amplitudes between each frame to integrate the overall distance. Key frames were selected, and the center position of the landmarks was used to compare with the gaze point of the participants matched by time point. The distance from participants' gaze point to the center position of the landmarks (deviation distance) was calculated and used as an indicator for the usage pattern of the perceptual landmarks along the simulated locomotion.

On one hand, a strong effect of distance was also found in the deviation distance that participants showed to better follow the location of the landmarks when the distance was short. This result should be treated with caution. There might be due to accuracy loss and sampling data loss for longer recording duration with dynamic stimuli. In our case, we were not able to rule out the influence from the device. On the other hand, we found the effect of interval condition that the deviation distance under the continuous condition was significantly shorter than the one under the static scene condition. In other words, when there were apparently blank between each static scene/frame, participants either did not try to imagine and follow the possible position of the landmarks, or they were not able to do so, which suggested that they did not use the motion-based mechanism. However, the reproduction performance showed participants were still able to reproduce the distance from only several static scenes with comparable accuracy as they did in the full visual cue condition. We could preliminarily believe that the distance estimation based on static mechanism functioned as well as the motion mechanism within the distance range used in the current experiment.

6 Implications and Future Work

Visual information is so far one of the most important senses in virtual reality, though we believe other senses have been incorporating gradually into the virtual world as technology progresses. Knowing how people perceive the space in a virtual world help to discover the cognitive process in real world and on the other hand, help to improve the sense of embodiment in the virtual world. It is well known that the space in virtual world is perceived compressed or less accurately compared to its counterpart of the real world [22, 23] when asking people to make a distance judgment. However, during simulated locomotion people were able to integrate the accurate distance from only a series of discrete static scenes at least in a certain range of distances. It implied that the estimation of the near space (within the action space in our case) is accurate and the integration process is well done too and the compression is probably caused by the perception of comparable farther space. Furthermore, it was suggested that the contribution of visual information in a full visual condition should not only be attributed to the effect of optic flow, especially when there were silent landmarks in the environment.

Our work made potential contributions for the future research in two aspects. First, we developed a paradigm to separate the static scenes from the optic flow in a simulated movement and made it possible to investigate the effect of only static scenes in locomotive spatial learning. Second, though we used a coarse analysis process for the eye movement data, it is suggested the possibility to use eye tracking technology on the studies of the dynamic stimuli or 3D space. Further work is needed to elaborate the data processing method so that we can fully use the visual information presented in the virtual reality and also match it more accurately with the eye movement data from the participants.

One limitation of our study is only the desktop virtual reality was used, and the results should be taken carefully because of the restricted field of view [24]. Further exploration with immersive virtual reality technology combined with eye tracking is needed to validate the findings in this study. In addition, in the current study, we attempted to explore the people's eye tracking behavior on the landmarks during the simulated movements, and we selected key frames from the dynamic stimuli to analyze where people were looking at along a simulated trip. However, only the gaze data from the learning phase were analyzed here. In the future we could also compare the eye tracking data from the reproduction phase and incorporate more index, saccades for example, to further unveil the integration process of this static mechanism.

Acknowledgments We would like to thank Dr. Frances Wang from University of Illinois at Urbana-Champaign and Dr. Wang Ying from Institute of Psychology, CAS for their instructions and constructive suggestions on our study. And we thank Dr. Yao Lin from Institute of Psychology, CAS for his help on data analysis. Special thanks are given to our participants for their patience and valuable time.

References

1. Warren WH (1995) Self-motion: visual perception and visual control. In: Epstein W, Rogers S (eds) *Perception of space and motion handbook of perception and cognition*. Academic Press, San Diego, pp 263–325
2. Klatzky RL, Loomis JM, Golledge RG, Cicinelli JG, Doherty S, Pellegrino JW (1990) Acquisition of route and survey knowledge in the absence of vision. *J Mot Behav* 22:19–43
3. Klatzky RL, Loomis JM, Golledge RG (1997) Encoding spatial representations through nonvisually guided locomotion: tests of human path integration. In: Medin D *The psychology of learning and motivation*, Academic Press, San Diego, pp 41–84
4. Berthoz A, Israël L, Georges-François P, Grasso R, Tsuzuku T (1995) Spatial memory of body linear displacement: what is being stored? *Science* 269:95–98
5. Lappe M, Frenz H, Bührmann T, Kolesnik M (2005) Virtual odometry from visual flow. *Proc SPIE* 5666:493–502
6. Redlick PF, Jenkin M, Harris RL (2001) Humans can use optic flow to estimate distance of travel. *Vision Res* 41:213–219
7. Wan X, Wang RF, Crowell JA (2012) The effect of landmarks in human path integration. *Acta Psychologica* 140(1):7–12
8. Ellmore TM, McNaughton BL (2004) Human path integration by optic flow. *Spat Cogn* 4(3):255–272
9. Gibson JJ (1950) *Perception of the visual world*. Houghton Mifflin, Boston
10. Kearns MJ, Warren WH, Duchon AP, Tarr MJ (2002) Path integration from optic flow and body senses in a homing task. *Perception* 31:349–374
11. Bremmer F, Lappe M (1999) The use of optical velocities for distance discrimination and reproduction during visually simulated self-motion. *Exp Brain Res* 127:33–42
12. Frenz H, Lappe M (2005) Absolute travel distance from optic flow. *Vision Res* 45:1679–1692
13. Frenz H, Lappe M, Kolesnik M, Bührmann T (2007) Estimation of travel distance from visual motion in virtual environments. *ACM Trans Appl Percept* 4(1):1–18
14. Riecke BE, van Veen HACH, Bühlhoff HH (2002) Visual homing is possible without landmarks: a path integration study in virtual reality. *Presence* 11(5):443–473
15. Lappe M, Jenkin M, Harris LR (2007) Travel distance estimation from visual motion by leaky path integration. *Exp Brain Res* 180:35–48
16. Loomis JM, Da Silva JA, Philbeck JW, Fukusima SS (1996) Visual perception of location and distance. *Curr Dir Psychol Sci* 5:72–77
17. Witt JK, Stefanucci JK, Riener CR, Proffitt DR (2007) Seeing beyond the target: environmental context affects distance perception. *Perception* 36:1752–1768
18. Sun H, Campos JL, Young M, Chan GSW (2004) The contributions of static visual cues, nonvisual cues, and optic flow in distance estimation. *Perception* 33:49–65
19. Frenz H, Lappe M (2006) Visual distance estimation in static compared to moving virtual scenes. *Span J Psychol* 9(2):321–331
20. Angelaki DE, Hess BJM (2005) Self-motion-induced eye movements: effects on visual acuity and navigation. *Nat Rev Neurosci* 6:966–976
21. Loomis JM, Klatzky RL, Golledge RG, Cicinelli JG, Pellegrino JW, Fry PA (1993) Nonvisual navigation by blind and sighted: assessment of path integration ability. *J Exp Psychol Gen* 122(1):73–91
22. Loomis JM, Knapp JM (2003) Visual perception of egocentric distance in real and virtual environments. In: Hettlinger LJ, Haas MW (eds) *Virtual and adaptive environments*. Erlbaum, Mahwah, pp 21–46
23. Thompson WB, Willemsen P, Gooch AA, Creem-Regehr SH, Loomis JM, Beall AC (2004) Does the quality of the computer graphics matter when judging distances in visually immersive environments? *Presence* 13:560–571
24. Riecke BE, Schulte-Pelkum J, Bühlhoff HH (2005) Perceiving simulated ego-motions in virtual reality—comparing large screen displays with HMDs. *Proc SPIE* 5666:344–355

Prediction Analysis of the Railway Track State Based on PCA-RBF Neural Network

Yan Yang, Xuyao Lu, Qi Dai and Hongjun Wang

Abstract With the development of high-speed railway, its security guarantee has received more and more attention. The railway track state is safety-critical and affected by many factors. The proposed approach focuses on establishing a track-state prediction model by monitoring data and analyzing the deformation trends in track geometric dimensions. Radial basis function (RBF) neural network is widely used in many industrial prediction domains, and principal component analysis (PCA) is a kind of methods to reduce dimensions. In this paper, we present an approach for predicting track irregularity index using PCA-RBF model. It is benefit for periodical maintenance of railway systems and safeguarding of transportation. The experiments show that the proposed model is effective.

Keywords Track-state prediction · Irregularity index · Radial basis function · Principal component analysis

1 Introduction

Railway track deformations may occur depending on different factors, including both internal factors (e.g., train and sleeper) and external factors (e.g., environment and human). Determining these deformations on time and taking precautions is

Y. Yang (✉) · X. Lu · Q. Dai · H. Wang

School of Information Science and Technology, Provincial Key Lab of Cloud Computing and Intelligent Technology, Southwest Jiaotong University, Chengdu, P.R. China
e-mail: yyang@swjtu.edu.cn

X. Lu

e-mail: lxy8228535@126.com

Q. Dai

e-mail: qdai@swjtu.edu.cn

H. Wang

e-mail: wanghongjun@swjtu.edu.cn

very important for the safety of railway systems [1, 2]. The track irregularity index is concrete embodiment of track structure deterioration and reflects the comprehensive performance of track state. Therefore, it is significant to study the variation characteristics of track irregularity and predict track deformations in future.

Yoshihiko proposed a track irregularity prediction method by regression method [3]. Canadian PWMIS (Railway Works Management Information System) prediction model [4] includes the track life prediction model, the database of model, the track quality state, and the track maintenance standards. Chang et al. [5] studied a multistage linear prediction model of track quality index (TQI). Qu et al. [6] investigated a track irregularity development prediction method based on Grey-Markov Chain. Gao [7] designed a track irregularity prediction method based on state transition probability matrix.

Radial basis function (RBF) neural network presented by Darken and Moody in 1989 [8] considers experience of evaluation expert and reduces uncertain of human difference, but cannot apply to track-state prediction directly due to prediction dynamic, track data relevance, and information overlapping. Principal component analysis (PCA) method can effectively reduce metadata relevance, noise, dimension, and net scale with keeping data information [9]. Thus, we introduce PCA to RBF to optimize input characters and improve performance of track-state prediction.

Based on geometric parameters and the application of PCA-RBF neural network, this paper proposes an approach to predict track state. Firstly, it selects features through PCA considering factors of track gauge, left alignment, right alignment, super elevation, left mean of heads (long or short) ALIGML/S, right mean of heads ALIGML/S, and TWIST [10]. Then, it obtains the biggest cumulative contribution rate as RBF input data. Finally, it trains and tests samples by RBF neural network and gets prediction value of railway track status.

The rest of the paper is organized as follows: In Sect. 2, we introduce the principal of PCA and RBF and outline the PCA-RBF model. Section 3 reports the experimental results and prediction analysis in detail. Section 4 provides conclusions and future work.

2 RCA-RBF Neural Network

PCA-RBF neural network model is composed of PCA and RBF. Through selecting the original data, PCA reduces the high-dimensional data to low-dimensional data. The factors with the higher cumulative contribution rate are chosen as the input value of the RBF, and then, the samples of these factors are trained and tested through the RBF neural network to get more accurate prediction value. Figure 1 shows the schematic diagram of PCA-RBF neural network model architecture.

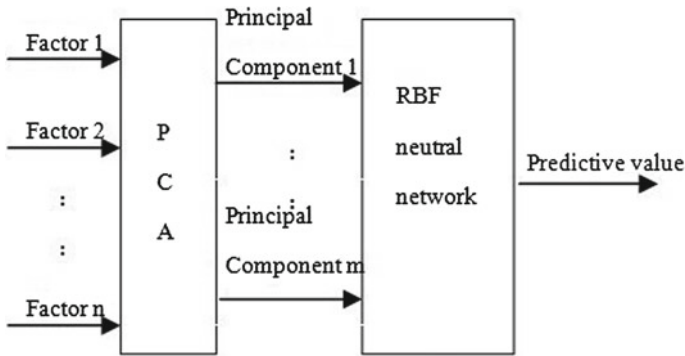


Fig. 1 Architecture for PCA-RBF neural network model

2.1 Principal Component Analysis

PCA is designed to simplify a number of relevant variables to a few unrelated comprehensive variables using the idea of dimension reduction, and these comprehensive variables contain the information of the original variables as much as possible, reducing the dimension of the data. PCA has the characteristic of keeping the data set to have the greatest contribution to difference. This is done by retaining the low-level principal component, ignoring the high-level-order principal components. So the low-level principal components can retain the most important aspect of the data [11]. The steps of PCA are outlined as follows:

Suppose X as an data table that has n samples and p variable, that is, $X = (x_{ij})_{n \times p}$, which $x_j = (x_{1j}, x_{2j}, \dots, x_{nj})^T \in R^n$ corresponds to the j .

- Standardizing the data, namely

$$\tilde{x} = \frac{x_{ij} - \bar{x}_j}{S_j} \quad (i = 1, 2, \dots, n; j = 1, 2, \dots, p). \tag{1}$$

where \bar{x}_j is the sample mean of the x_j, S_j is the sample standard differential of the x_j .

- Calculating the covariance matrix V of the standardized data matrix X . Then, V is the relationship matrix of X .
- Calculating the first m feature values of $V: \lambda_1 \geq \lambda_2 \geq \dots \lambda_m$, and the corresponding unit feature vectors $a_1, a_2, \dots a_m$, requiring that they are standard orthogonal.

$$\begin{aligned} a_1 &= (a_{11}, a_{21}, \dots a_{n1})^T \\ a_2 &= (a_{12}, a_{22}, \dots a_{n2})^T \\ &\dots\dots\dots \\ a_n &= (a_{1n}, a_{2n}, \dots a_{nn})^T. \end{aligned} \tag{2}$$

- Calculating the principal component h

$$x_h' = a_{h1}x_1 + a_{h2}x_2 + \cdots + a_{hp}x_p. \quad (3)$$

- Calculating the cumulative contribution rate of the principal component m

$$Q_m = \frac{\sum_{h=1}^m \lambda_h}{\sum_{j=1}^p S_j^2}. \quad (4)$$

When $Q_m \geq 85\%$, principal component analysis is completed. We select the m principal components from these principal components as the input of the RBF neural network.

2.2 RBF Neural Network

RBF neural network has time series prediction ability. The work reported in this paper uses RBF neural network due to its structural simplicity and prediction capabilities. RBF neural network consists of an input layer, a hidden layer with a nonlinear RBF activation function, and a linear output layer. Figure 2 shows the structure of the RBF [12].

where, the input layer has m -dimensional vector $x' = \{x_1', x_2', \dots, x_m'\}$ through the PCA as the input of the RBF neural network. The hidden layer is the l -dimensional vector $R = \{R_1, R_2, \dots, R_l\}$, and the function is the nonlinear transformation of the input layer information. The number of hidden layer nodes is determined by training; here, we use Gaussian kernel function [13], the formula is as follows:

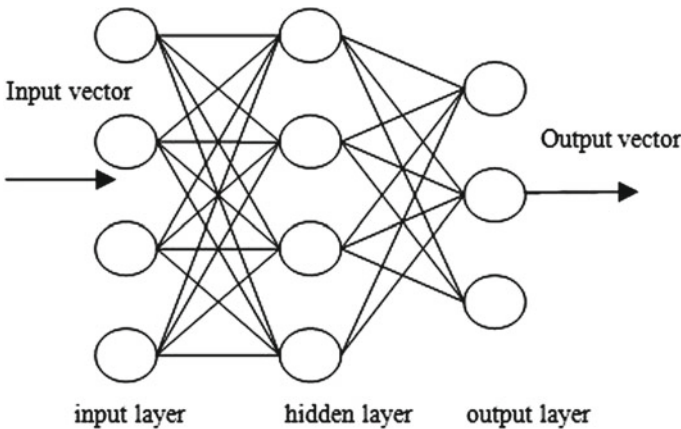


Fig. 2 Structure of the RBF

$$R_i(X') = \exp \left[-\frac{(X' - c_i)^T (X' - c_i)}{2\sigma^2} \right] \quad (i = 1, 2, \dots, l). \tag{5}$$

where $R_i(X)$ is the output of the i th hidden layer node, c_i is the center value of the Gaussian function, σ_i is the variance of Gaussian function, l is the number of nodes of hidden layer. So we need to learn the c_i , σ_i and l . The output layer is generally a simple linear function through the transformation of the hidden layer:

$$f(X') = \sum_{i=1}^l W_i R_i(X'). \tag{6}$$

where W_i is the weight from the i th hidden layer node to the output unit, generally use the least squares method to estimate.

2.3 Description of the PCA-RBF Model

The pseudo-code of the PCA-RBF model is shown in Algorithm 1. The input of model is the original value of the data set, and the output is the predicted value of the data. Where Q_m is the cumulative contribution rate of M principal components; V is the covariance matrix of data X ; λ_m and a_m are the feature values and the feature vectors of V ; x_h' is the h th principal component; and W_i is the weights from the i th hidden layer node to the output unit.

Algorithm 1: Pseudo-code of the PCA-RBF model

Input: Data X

Output: Predicted data

begin

Standardize X according to the formula (1);

if $Q_m \geq 85\%$ **then**

Calculate covariance matrix V ;

Calculate λ_m and a_m ;

Calculate X_h' according to the formula (3);

Calculate Q_m according to the formula (4);

end

Calculate the nodes of hidden layer according to the formula (5);

Estimate W_i according to the OLS;

Calculate X' according to the formula (6)

end

3 Experimental Analysis

3.1 Data Set

The performance of the proposed method has been evaluated on multiple real data sets. The evaluation index is relative error and TQI between prediction and original value [5, 14]

$$\text{TQI} = \sum_{i=1}^7 \sigma_i. \quad (7)$$

where $\sigma_i = \sqrt{\frac{1}{n} \sum_{j=1}^n (x_{ij}^2 - \bar{x}_i^2)}$, σ_i is the standard differential of the each geometric deviation, $\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ij}$, $i = 1, 2, \dots, 7$, are the influencing factors, including track gauge, left alignment, right alignment, super elevation, left mean of heads (long or short) ALIGML/S, right mean of heads ALIGML/S, and TWIST, x_{ij} is the amplitude of the geometric deviations in the 200 m unit section, $j = 1, 2, \dots, n$, $i = 1, 2, \dots, 7$, n is the number of sampling points $n = 800$ in the 200 m unit section.

3.2 Experimental Results

The first six batches of data of the line are as a training set, and the seventh batch of data is as the test set. Firstly, the seven factors (track gauge, left alignment, right alignment, super elevation, left mean of heads (long or short) ALIGML/S, right mean of heads ALIGML/S, and TWIST) are reduced by PCA. Then, three factors which the cumulative contribution rate is greater than 85 % are as the input of the RBF neural network. Finally, the prediction values are obtained by the RBF network. A part of results is shown in Table 1. Where GONGLI and MI are the distance of the measuring point, P_GUIJU, P_CHAOGAO, P_ZGX, P_YGX, P_ZGD, P_YGD, and P_SJK are the predicted values of track gauge, super elevation, left alignment, right alignment, left mean of heads (long or short) ALIGML/S, right mean of heads ALIGML/S, and TWIST of the seventh batch.

The relative error = This work is supported - original value/original value. The relative error of the Table 1 is shown in Table 2. Where E_GUIJU, E_CHAOGAO, E_ZGX, E_YGX, E_ZGD, E_YGD, and E_SJK are the relative errors between the original value and predicted value of track gauge, super elevation, left alignment, right alignment, left mean of heads (long or short) ALIGML/S, right mean of heads ALIGML/S, and TWIST. Table 3 lists the TQI values of the original value and predicted values, where P_TQI is the predicted value of TQI, and E_TQI is the relative error.

Table 1 The original values and predicted values of seven impact factors

GONGLI (km)	MI (m)	GUIJU	CHAOGAO	ZGX	YGX	ZGD	YGD	SJK
1239	810.5	-0.58	-2.41	-0.69	-0.57	-0.64	-0.08	0.86
1239	810.75	-0.59	-2.13	-0.85	-0.73	-0.65	-0.2	1.02
1239	811	-0.49	-1.82	-0.76	-0.74	-0.59	-0.34	0.86
1239	811.25	-0.5	-1.86	-0.71	-0.7	-0.55	-0.47	1.05
1239	811.5	-0.3	-1.58	-0.64	-0.84	-0.5	-0.55	0.89
GONGLI (km)	MI (m)	P_GUIJU	P_CHAOGAO	P_ZGX	P_YGX	P_ZGD	P_YGD	P_SJK
1239	810.5	-0.66	-1.44	-0.77	-0.76	-0.62	-0.06	0.56
1239	810.75	-0.67	-1.48	-0.79	-0.82	-0.66	-0.16	0.82
1239	811	-0.61	-1.29	-0.76	-0.79	-0.64	-0.18	0.69
1239	811.25	-0.63	-1.37	-0.51	-0.63	-0.3	-0.25	1.06
1239	811.5	-0.54	-1.2	-0.48	-0.63	-0.3	-0.25	0.92

Table 2 Relative error between original value and predicted value of Table 1

GONGLI(km)	MI(m)	E_GUIJU	E_CHAOGAO	E_ZGX	E_YGX	E_ZGD	E_YGD	E_SJK
1239	810.5	0.12	-0.40	0.12	0.33	-0.03	-0.25	-0.35
1239	810.75	0.14	-0.31	-0.07	0.12	0.02	-0.2	-0.2
1239	811	0.24	-0.29	0.00	0.07	0.08	-0.47	-0.2
1239	811.25	0.26	-0.26	-0.28	-0.10	-0.31	-0.47	0.01
1239	811.5	0.80	-0.24	-0.25	-0.25	-0.4	-0.55	0.03

Table 3 Comparison of TQI value between original value and predictive value in 1,239 Km

Distance segment (km)	TQI	P_TQI	E_TQI
1239-1239.2	1.49	1.23	0.18
1239.2-1239.4	2.35	2.50	-0.06
1239.4-1239.6	2.28	2.33	-0.02
1239.6-1239.8	2.59	2.79	-0.08
1239.8-1240	2.49	2.18	0.12

In order to compare the original value and predicted value intuitively, Figs. 3, 4, 5, 6, 7, 8, 9 show the original value and predicted value of the track gauge, super elevation, left alignment, right alignment, left mean of heads (long or short) ALIGML/S, right mean of heads ALIGML/S, and TWIST, respectively, where the horizontal axis denotes distance (unit: km), and each vertical axis denotes seven factors, respectively. The solid blue line “×” indicates the original value and the solid red line “·” indicates the predicted value. Figure 10 shows the original value and predicted value of the 50 TQI value in a distance. Where the horizontal axis denotes data point, namely the interval of each data point is 200 m. The vertical axis denotes the TQI. The solid blue line “×” indicates the original value, and the solid red line “·” indicates the predicted value.

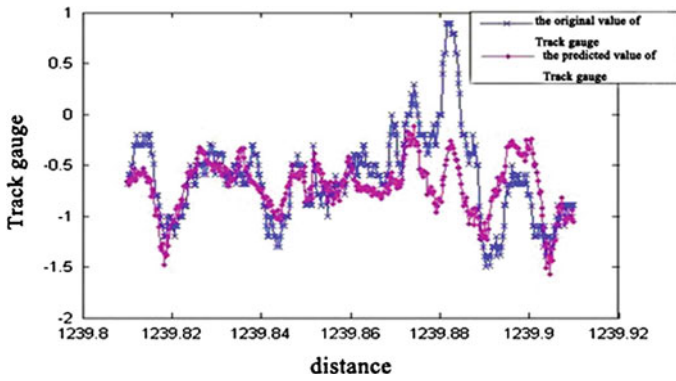


Fig. 3 Track gauge

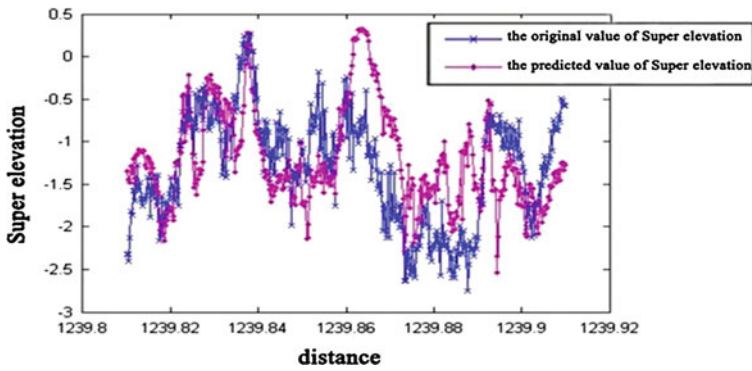


Fig. 4 Super elevation

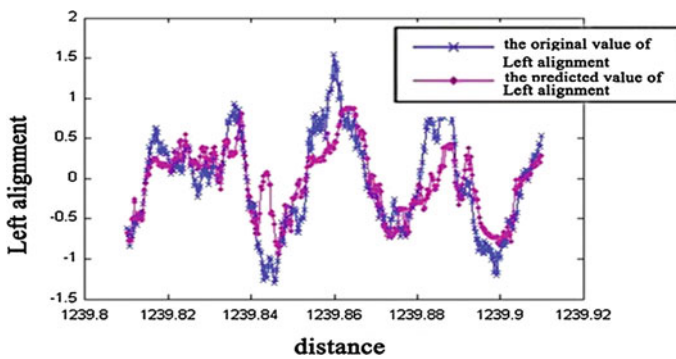


Fig. 5 Left alignment

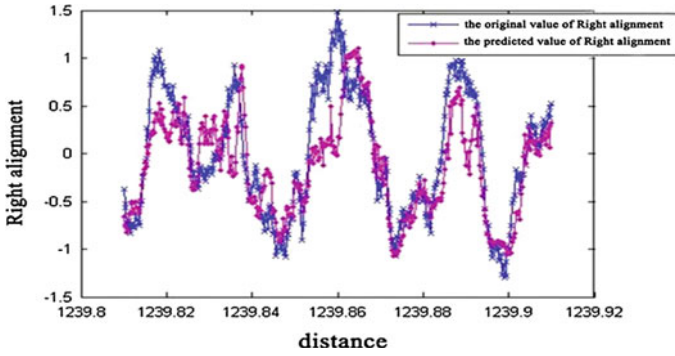


Fig. 6 Right alignment

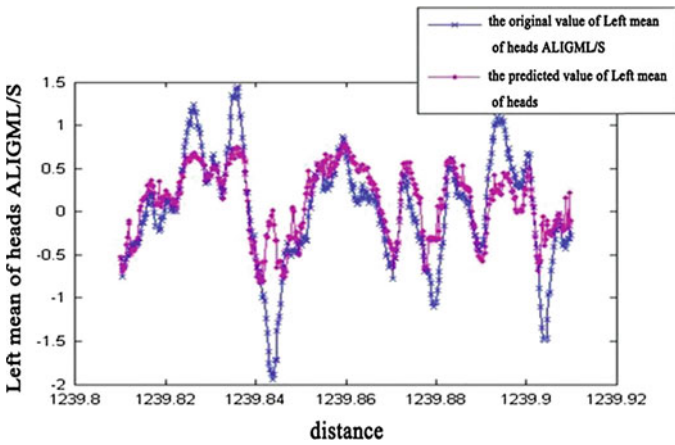


Fig. 7 Left mean of heads (long or short) ALIGML/S

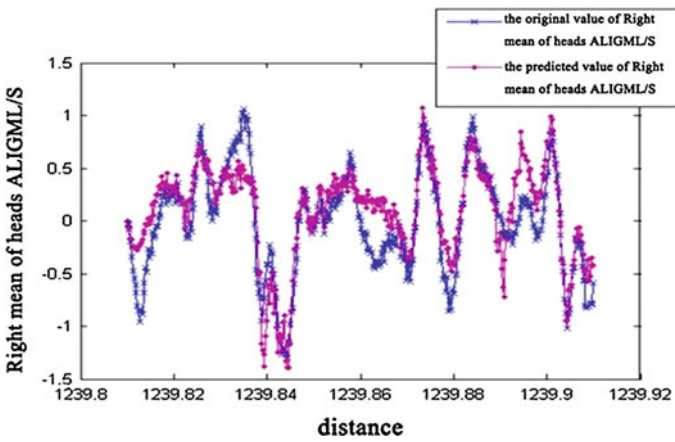


Fig. 8 Right mean of heads (long or short) ALIGML/S

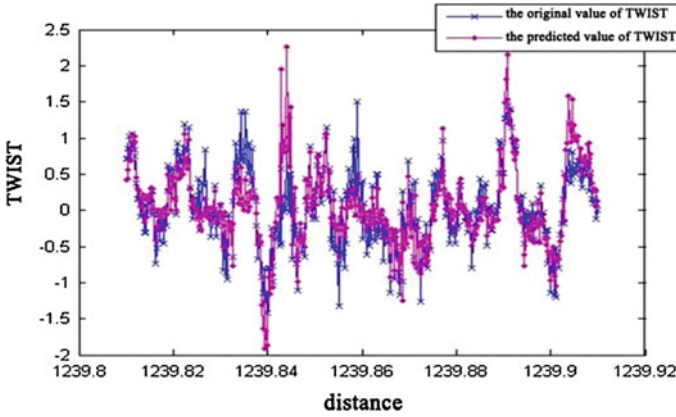


Fig. 9 TWIST

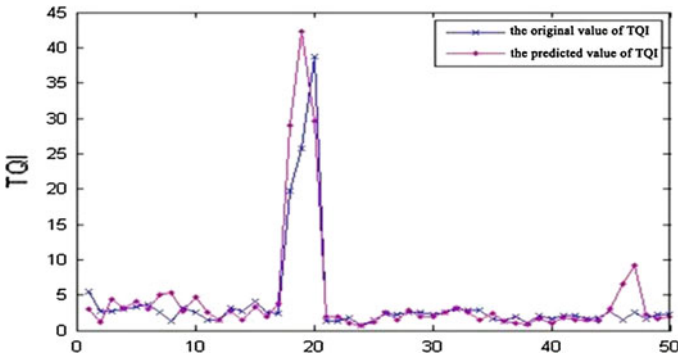


Fig. 10 TQI

4 Conclusions

In this paper, we predicted railway track irregularity index using a PCA-RBF neural network model. Our work may reduce dimension effectively, eliminate the relativity of all impact factors, and get a better TQI predicted value. We note that the predicted value is almost the same as original value except mean of heads (long or short) ALIGML/S. The reason may be mutual impact of each factor. In extreme conditions, such as train departure, destination, curve, and tunnel section, the relative error between prediction and original value is increased, and TQI value is also bigger than normal value. For future work, we will consider more factors, such as curvature, horizontal acceleration, vertical acceleration to improve prediction accuracy.

Acknowledgments This work is supported by the National Science Foundation of China (Nos. 61134002 and 61170111), the Science and Technology Research Funds of the Ministry of Railways (2010G006-D), and Research Funds of Traction Power State Key Laboratory of Southwest Jiaotong University (2012TPL_T15)

References

1. Luo L, Zhang G, Wu W (2006) Control of the track irregularity in the wheel-rail system. China Railway Publishing House, Beijing
2. Akpinar B, Glal E (2012) Multisensor railway track geometry surveying system. *IEEE Trans Instrum Meas* 61(1):190–197
3. Yoshihiko S (2001) New track mechanics. China Railway Publishing House, Beijing
4. Huang Y (2009) Study on prediction method for railway track irregularity. Master Dissertation, Beijing Jiaotong University
5. Chang H, Liu R, Wang W (2010) Multistage linear prediction model of track quality index. In: *Proceeding the conference on traffic and transportation studies*. ICTTS, Kunming, pp 1183–1192
6. Qu J, Gao L, Xin T (2010) Track irregularity development prediction method based on Grey-Markov chain model. *J Beijing Jiaotong Univ* 34(4):107–111
7. Gao J (2011) Track irregularity development prediction research based on state transition probability matrix. *Railway Constr* 7:140–143
8. Moody J, Darken CJ (1989) Fast learning in networks of locally-tuned processing units. *Neural Comput*, pp 281–294
9. Liu Y, Huang D, Li Y (2011) Adaptive statistic process monitoring with a modified PCA. In: *Proceedings IEEE international conference on computer science and automation engineering (CSAE2011)*. IEEE Press, Shanghai, pp 713–716
10. Chen D, Tian X (2008) Detection technology development of the China’s high-speed railway track. *Railway Constr* 12:82–86
11. Sing JK, Thakur S, Basu DK, Nasipuri M (2008) Direct kernel PCA with RBF neural networks for face recognition. In: *Proceedings IEEE region 10 annual international conference*. IEEE Press, Hyderabad, pp 1–6
12. Salahshoor K, Kordestani M, Khoshro MS (2009) Design of online soft sensors based on combined adaptive PCA and RBF neural networks. In: *Proceedings IEEE symposium on computational intelligence in control and automation (CICA2009)*. IEEE Press, Nashville, TN, pp 89–95
13. Wei H, Amari S (2007) Eigenvalue analysis on singularity in RBF networks. In: *Proceedings IEEE international conference on neural networks*. IEEE Press, Orlando, FL, pp 690–695
14. Xu P, Sun Q, Liu R, Wang F (2011) A short-range prediction model for track quality index. *Proc Inst Mech Eng, Part F: J Rail Rapid Transit* 225(3):277–285

A Two-Step Agglomerative Hierarchical Clustering Method for Patent Time-Dependent Data

Hongshu Chen, Guangquan Zhang, Jie Lu and Donghua Zhu

Abstract Patent data have time-dependent property and also semantic attributes. Technology clustering based on patent time-dependent data processed by trend analysis has been used to help technology relationship identification. However, the raw patent data carry more features than processed data. This paper aims to develop a new methodology to cluster patent frequency data based on its time-related properties. To handle time-dependent attributes of patent data, this study first compares it with typical time series data to propose preferable similarity measurement approach. It then presents a two-step agglomerative hierarchical technology clustering method to cluster original patent time-dependent data directly. Finally, a case study using communication-related patents is given to illustrate the clustering method.

Keywords Patent analysis · Technology clustering · Patent time-dependent data · Agglomerative hierarchical clustering

H. Chen (✉) · G. Zhang · J. Lu
Decision Systems and e-Service Intelligence Lab, Centre for Quantum Computation
and Intelligent Systems, Faculty of Engineering and Information Technology,
University of Technology, Sydney, Australia
e-mail: hongshu.chen@student.uts.edu.au

G. Zhang
e-mail: guangquan.zhang@uts.edu.au

J. Lu
e-mail: jie.lu@uts.edu.au

H. Chen · D. Zhu
School of Management and Economics, Beijing Institute of Technology, Beijing, China
e-mail: zhudh111@bit.edu.cn

1 Introduction

Patent is the main manifestation of intellectual property for a company or a government sector. The study of using patents as indicators for technology analysis has begun from early 1980s, since then patent analysis became a very useful tool that utilized to support technology research and development (R&D) planning, competition analyses, and analytic studies of how technologies emerge, mature, and disappear [1]. Under recent circumstances of fast and complex technological advances, data mining for patent analysis has become increasingly important to decision-making of both private companies and governments for continuous technology development and risk reducing.

In general, previous study of patent analysis focuses on monitoring technological horizon, forecasting the future trend [2], identifying emerging technology [3, 4], clustering and classifying technologies [5], and constructing technology intelligence systems [6]. When doing research, the patent database of United States Patent and Trademark Office (USPTO) is mainly used. This is because patents submitted in other countries are often simultaneously submitted in United States, which making USPTO database the most representative and standard system for technology analyzing. In the system, each patent will be categorized by both international patent classification (IPC) and United States patent classification (USPC) when it is applied. Since IPC and USPC are based on predetermined technological boundary, although sharing keywords in a same technical area, patents within related IPC and USPC have shown different growth patterns with varied technological innovation activities. To investigate the natural development of various technologies within a certain industry field, we need to cluster the patents and estimate their actual relationships also group patterns, which make technology clustering one of the important parts of patent analysis.

Technology clustering helps with promoting the further examination of possible inter-relationships between technologies in IPC or USPC system (e.g., substitutive or complementary) and thus makes it possible to identify technologies that are in the same levels of their life cycles [7]. Patents have both time-dependent property and semantic attributes. Matching with two types of properties, there are basically two ways to perform technology clustering: patent trend-based clustering methods [7] and content-based clustering techniques [8]. Patent content clustering is used to cluster different patent documents into homogenous groups, so that semantic relationship between patents is identified. In this research, we mainly focus on the former case, trend-based technology clustering. For a particular industry area, patents applying and granting frequency during a period of time can be presented as several sequences of time-dependent data. In previous study, Hyoung-joo and Sungjoo [9] used Hidden Markov Model to analyze growth patterns of patent frequency data and then cluster the technologies with the trend patterns. Heuristics methods are also used to deal with technology clustering problem by using patents count on base of ICP [10].

Focusing on time-related properties of patents, the aims of this study is to explain how to use patent frequency sequence data to cluster related technologies directly. We will first analyze the properties of patent time-related data in details, and compare it with typical time series data and then discuss how to use traditional measurement of time series into research of patent time-dependent data and whether time-dependent data should be normalized or not before clustering. Treating patent annual grant number of a major class from USPC as a vector, this research will then present a two-step agglomerative hierarchical technology clustering method to cluster technologies with original data which has lossless trend features. Finally, communication-related patents will be analyzed using two-step clustering method as a case study.

The remainder of this paper is organized as follows. [Section 2](#) analyzes the characteristics of patent time-dependent data and illustrates the methodology. In [Sect. 3](#), we propose a two-step agglomerative hierarchical technology clustering method toward patent time-dependent sequence data. [Section 4](#) presents a case study of communication-related technology to validate the effectiveness of the proposed clustering method. Final section concludes the paper and expounds the future study.

2 Patent Data Analysis and Methodology

In this section, we will first analyze the properties of patent time-dependent data in details and then present the methodology of this research. Whether Euclidean distance normalization can be used or not while dealing with patent frequency data will also be discussed.

2.1 Patent Time-Dependent Data

Within a particular industry area, annual counts of each relevant USPC or IPC during a period of time can be presented as one vector. Here, since we only focus on the annual number of each major class, the data are discrete. Different with voice signal data, stock market price data and other typical time series data, for patents frequency data sequences, there are much less “recognizable repeated patterns” which shown as swing of the data. Technology development and increased interest in intellectual property protection have both promoted the growth of patent application and grant. Therefore, the time-dependent data of patents annual number mostly show an upward tendency or remains stable. Moreover, because the issue number for each year is fixed, patent frequency data have highly strong relationship with time, which makes classical approaches of similarity measures like dynamic time warping [11] using in time series analysis are not suitable for this case. When dealing with patent time-dependent data, we

are not able to “warp” the time axis of any patent frequency series but have to accept all unaffected difference between two vectors. When dealing with similarity measurements of patents time-dependent sequences, Euclidean distance is the traditional but effective method.

In this study, we choose annual counts of issued patents to create time-dependent data sequences. Let $P_i = \{p_{i1}, \dots, p_{in}\}$ defines one sequence, here i indicates the number of USPC classes within target technology area, while n stands for the number of years.

2.2 Agglomerative Hierarchical Clustering

Hierarchical clustering is a data-mining method which focuses on building a hierarchy of clusters. There are two types of hierarchical clustering methods: agglomerative and divisive depending upon whether a “bottom up” or “top down” approach is followed [12]. Agglomerative hierarchical clustering (AHC) merges the atomic cluster, which only contain single object itself, into larger clusters until all the objects are in a big cluster [13]. Here, the Ward’s linkage approach is most widely used when merging the clusters [14].

$$d(p, q) = \sqrt{\frac{2n_p n_q}{(n_p + n_q)} \|\bar{x}_p - \bar{x}_q\|_2} \quad (1)$$

where

- $\|\bar{x}_p - \bar{x}_q\|_2$ is Euclidean distance
- \bar{x}_p and \bar{x}_q are the centroids of clusters p and q
- n_p and n_q are the number of elements in clusters p and q

2.3 Could Euclidean Distance Normalization be Used to Preprocess the Patent Sequences?

One major disadvantage of Euclidean distance is that it is brittle [15]. This problem can be solved by normalizing the sequences before calculating the distance. Goldin and Kanellakis [16] described a normalization approach for improving effect of Euclidean distance computing. As mentioned $P = \{p_1, \dots, p_n\}$ stands for a sequence, $\mu(P)$ is the mean of P and $\sigma(P)$ is the standard deviation of P . The normalized sequence P' can be calculated as follows:

$$P'_i = \frac{p_i - \mu(P)}{\sigma(P)} \quad (2)$$

As mentioned above, patent sequences are not typical time series data. Then, could Euclidean distance normalization be used here to preprocess the data? For instance, two sequences A and B have approximately the similar shape indicate that they have roughly the same growth pattern. In some cases, A and B may have different offsets in the Y -Axis, and Euclidean distance between these two sequences will still be relatively large although they have basically the same growth pattern [17]. This is the reason that traditional time series data must be normalized before being clustered. However, when dealing with patent frequency data, the offsets of sequences define the issue amount at the initial observation year, and the different Y -axis level of the data sequences indicates that they have dissimilar amount of patent issue during the observation years. Here, amount of the patents is also a pattern of the data sequences. That is, different with traditional time series data, we are not able to simply “warp” the Y -axis to get better trend similarity comparison.

After understanding the requirement of analyzing patent time-dependent data in its entirety, we think that more accurate comparison between different sequences based on trend features which needs data normalization is undoubtedly wanted, and at the same time, we also require to keep the “amount patterns” which represented by different Y -axis levels. To solve this problem, we consider the clustering in two steps.

3 Two-Step Agglomerative Hierarchical Clustering Method for Patent Time-Dependent Data

“Trend” and “amount” are both patterns of patent time-dependent data sequences. “Trend” pattern is more related to developing features of one technology, which means that technologies with similar trends will appear in approximately the same technological life cycle stage. “Amount” pattern thus indicates the existing developed feature of the target technology, which explains the development popularity. Based on the characteristics of patent time-dependent data, we present two-step AHC method to cluster technologies using raw data from USPTO database. The overall process of two-step AHC is shown as Fig. 1, and the steps are as follows:

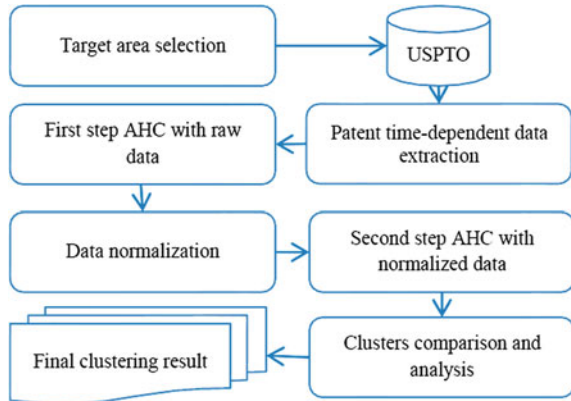
Step 1: A target technology area needs to be selected according to demands of users (decision-makers of private companies or governments) and mapped to several related USPC.

Step 2: We then query relevant USPC in USPTO database and get their grant amount of each year to form several data sequences.

Step 3: After data extraction, we employ the first step AHC to cluster the raw data, which keeps the “amount” pattern of each data sequence.

Step 4: Then, the clustering result will be recorded, and the data will be normalized subsequently.

Fig. 1 Process of two-step agglomerative hierarchical clustering for patent time-dependent data



Step 5: For preferable “trend” pattern comparison, after normalizing the data sequences, we proceed to do the second time of AHC with normalized data.

Step 6: Clusters comparison and analysis are presented to discuss the final clustering result.

Step 7: Finally, we consider the intersection of two steps clustering produces the final technology groups, which reflects both “trend” patterns and “amount” differences of the technologies.

4 Result and Discussion

A case study is performed in this section using communication-related patents to explain how the two-step AHC method works.

4.1 Data

Based on official document of “Classes within the U.S. Classification System” published on USPTO website [18], we identified 8 communication-related technology classes as shown in Table 1. For the 8 classes identified, the annual number of patents issued from 1963 to 2011 in these USPC were collected and presented by 8 data sequences. Each data sequence stands for development trend of one technology that corresponds to a USPC class with 49 elements. The numbers of patents issued in each technology are shown in Fig. 2. Although the data sequences swing up and down during the first 20 years, the whole trends of relevant technologies revealed an increasing tendency as a whole.

To prepare for the next few research steps, we then normalized 8 patent data sequences using the approach mentioned above. The data after normalization are

Table 1 Communication-related USPC

Class number	Name of the patent class
340	Communications: electrical
342	Communications: directive radio wave systems and devices
343	Communications: radio wave antennas
367	Communications, electrical: acoustic wave systems and devices
370	Multiplex communications
375	Pulse or digital communications
379	Telephonic communications
455	Telecommunications

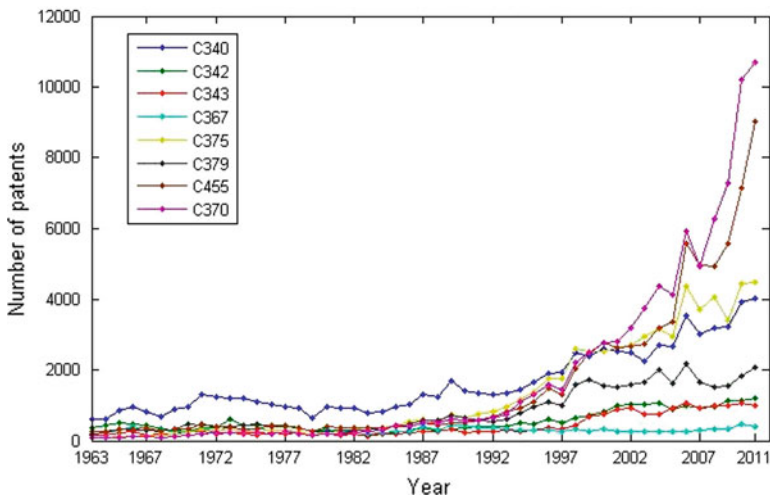


Fig. 2 Patent time-dependent data of communication-related technology from USPC

shown in Fig. 3. Compare to Fig. 2 that showing all the patent trends rising regardless of the technology type, the trend of technology 367 after normalization appears with strong fluctuations, which significantly different with other sequences.

4.2 Two-Step AHC Technology Clustering

We measure the similarity of two technologies by computing the distance between their vectors using Euclidean distance method. Then, AHC approach with Ward’s linkage algorithm is utilized to the distance measurement result, a 8×8 distance matrix, to cluster the technologies.

Dendrogram produced by the first step AHC using raw patent time-dependent sequences data is shown in Fig. 4, where technologies with closer distance are

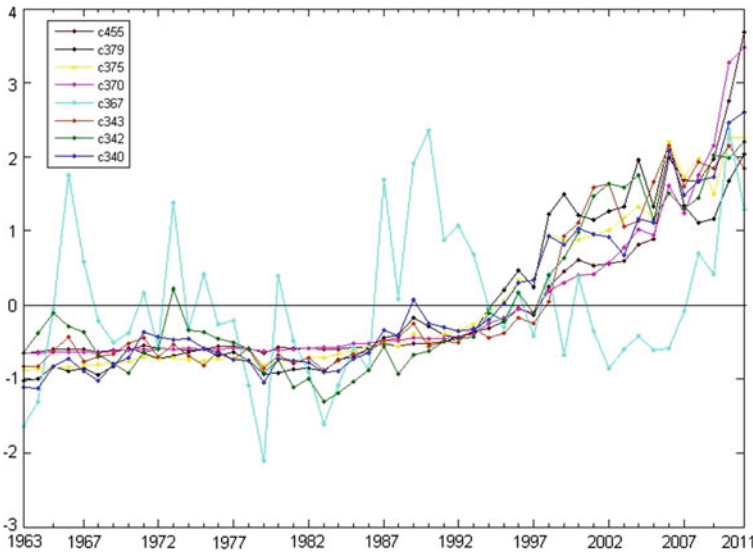
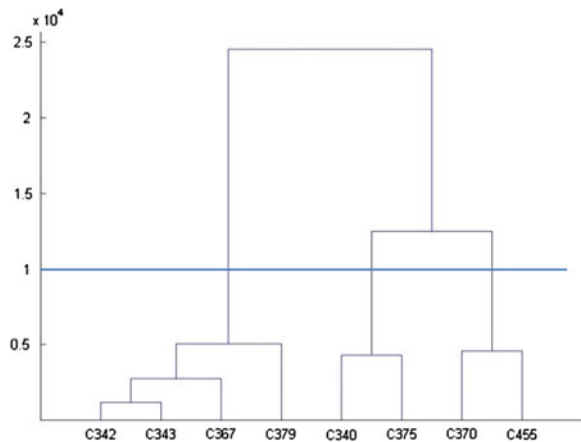


Fig. 3 Patent time-dependent data of communication-related technology after normalization

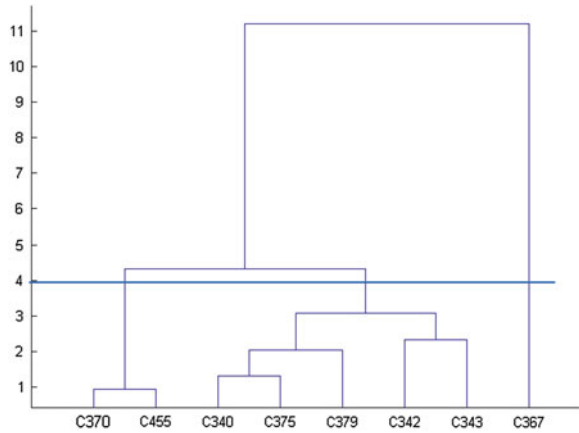
Fig. 4 Clustering result of the first step AHC



linked together. If we cluster 8 technologies into three main groups, we can see from dendrogram directly that Cluster 1 includes technologies 342, 343, 367 and 379, where 342 and 343 are even closer in the group; Cluster 2 includes technologies 340 and 375; Cluster 3 includes technologies 370 and 455. Here, we use the horizontal line in Fig. 4 to locate where we have chosen the number of groups. This result still takes the patent amount into clustering consideration, which showing that technology 342 and 343 are most similar to each other.

Dendrogram produced by the second step AHC using normalized patent time-dependent sequences data is shown in Fig. 5. After the normalization, the trend

Fig. 5 Clustering result of the second step AHC



pattern of each sequence is amplified. If we still cluster 8 technologies into three main groups, the result is listed as follow: Cluster 1 includes technologies 370, 455; Cluster 2 includes technologies 340, 375, 379, 343 and 343, where technologies 340, 375 and 379 have more similar trend patterns, technologies 342 and 343 share more similarities; Cluster 3 only contains technology 367, which means it is an outlier if we mainly consider the trend patterns. The horizontal line in Fig. 5 locate where we have chosen the number of groups.

4.3 Result Comparison and Discussion

In this research, we identify the intersection of two steps clustering results as the final technology groups, which reflects both trend patterns and amount differences of the technologies.

Figure 6 shows the comparison of clustering results after each step of AHC. We can observe the final groups as follow, Cluster 1: technologies 342

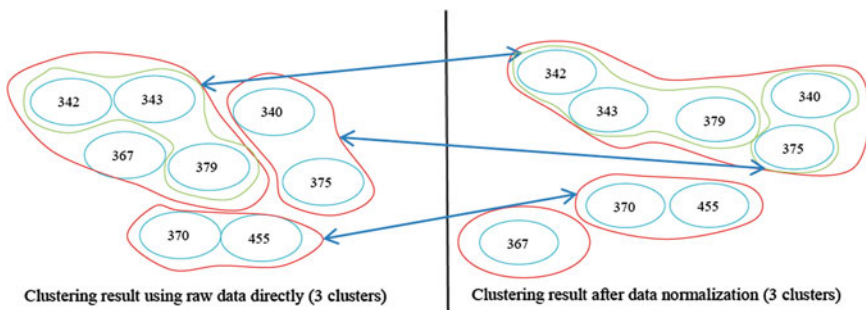


Fig. 6 Clustering result comparison of two steps AHC

(Communications: Directive Radio Wave Systems and Devices), 343 (Communications: Radio Wave Antennas) and 379 (Telephonic Communications); Cluster 2: technologies 340 (Communications: Electrical) and 375 (Pulse or Digital Communications); Cluster 3: technologies 370 (Multiplex Communications) and 455 (Telecommunications). Technology 376 is an outlier in this case study, which can be testified in the original data plot Fig. 2. It shows obviously less rising trend and lower Y -axis than other technologies. Moreover, no matter from the angle of trend correspondence or amount similarity, technology 370 and technology 455 are more similar, and technologies 342, 343, and 379 share more of the same features, and technologies 340 and 375 can be seen as a group.

5 Conclusion and Future Study

Modern society emphasizes the role of technology R&D increasingly, which promotes technical advances and accumulation of intellectual property at the same time. Technology clustering helps with analyzing the inner-relationship between different technologies within a same industry to optimize technology management and evaluation process, in both the public and private domains. This paper discussed how to measure similarity of patent frequency sequences data preferably and how to use it to cluster related technologies directly. By discussing the time-related properties of patents, we concluded that compared with typical time series data, patent time-dependent data has both “trend” and “amount” features; thus, we are not able to warp X -axis or Y -axis to process complicated similarity measurement, which makes Euclidean distance a traditional but effective method to solve the problem. To cluster technologies more efficiently and accurately, we then present a two-step agglomerative hierarchical technology clustering method to group original patent time-dependent data directly. A case study using communication-related patents data was then given to illustrate the method.

In conclusion, this method will be useful when we only need to identify technology cluster within a particular industry area. Raw data carry more features than processed data. Patterns of the patent time-dependent data sequences will be more easily extracted by using the original data. Moreover, this method can be utilized when we need to identify the similar technologies with existing ones or we want to exclude outliers. Finally, clustering the target technologies before trend identification will make trend analysis more efficiently.

Although this method is able to cluster technologies with raw data and takes consideration of both “trend” and “amount” features, it still needs to work with technology trend analysis to access more comprehensive and meaningful result. Therefore, the future study will focus on analyzing and forecasting technology trend after clustering process. What is more, semantic attributes of patent data also need to be taken into consideration, which makes it possible to analyze patent data in angles of both semantics and trend at the same time; thus, we can access multidimensional knowledge of technology development in a certain industry area.

References

1. Richard SC (1983) Patent trends as a technological forecasting tool. *World Pat Inf* 5(3):137–143
2. Cozzens S et al (2010) Emerging technologies: quantitative identification and measurement. *Technol Anal Strateg Manag* 22(3):361–376
3. Bengisu M, Nekhili R (2006) Forecasting emerging technologies with the aid of science and technology databases. *Technol Forecast Soc Chang* 73(7):835–844
4. Robinson DKR et al (2013) Forecasting innovation pathways (FIP) for new and emerging science and technologies. *Technol Forecast and Soc Chang* 80(2):267–285
5. Chen Y-L, Chang Y-C (2012) A three-phase method for patent classification. *Inf Process & Manag* 48(6):1017–1030
6. Yoon J, Kim K (2012) TrendPerceptor: a property-function based technology intelligence system for identifying technology trends from patents. *Expert Syst Appl* 39(3):2927–2938
7. Lee H, Lee S, Yoon B (2011) Technology clustering based on evolutionary patterns: the case of information and communications technologies. *Technol Forecast Soc Chang* 78(6):953–967
8. Trappey CV et al (2011) Using patent data for technology forecasting: China RFID patent analysis. *Adv Eng Inform* 25(1):53–64
9. Lee S, Lee H, Yoon B (2012) Modeling and analyzing technology innovation in the energy sector: patent-based HMM approach. *Comput & Ind Eng* 63(3):564–577
10. Dereli T et al (2011) Enhancing technology clustering through heuristics by using patent counts. *Expert Syst Appl* 38(12):15383–15391
11. Berndt D, Clifford J (1994) Using dynamic time warping to find patterns in time series. In: *Workshop on knowledge discovery in databases KDD-94 proceedings, Seattle*
12. Warren Liao T (2005) Clustering of time series data—a survey. *Pattern Recognit* 38(11):1857–1874
13. Keogh E, Pazzani M (1998) An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback. In: *Workshop on knowledge discovery in databases KDD-98 proceedings*
14. Ward JH (1993) Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* 58:236–244
15. Maimon O, Rokach L (2010) *Data mining and knowledge discovery handbook, vol 1*. Springer Science + Business Media, LLC
16. Goldin D, Kanellakis P (1995) On similarity queries for time-series data: constraint specification and implementation. In: *The 1st international conference on the principles and practice of constraint programming, Springer, Cassis*
17. Keogh E, Kasetty S (2003) On the need for time series data mining benchmarks: a survey and empirical demonstration. *Data Min Knowl Disc* 7(4):349–371
18. United States Patent and Trademark Office, Classes within the U.S. Classification System (2012). <http://www.uspto.gov/patents/resources/classification/classescombined.pdf>

A Novel Modular Recurrent Wavelet Neural Network and Its Application to Nonlinear System Identification

Haiquan Zhao and Xiangping Zeng

Abstract To reduce the computational complexity and improve the performance of the recurrent wavelet neural network (RWNN), a novel modular recurrent neural network based on the pipelined architecture (PRWNN) with low computational complexity is presented in this paper. Its modified adaptive real-time recurrent learning (RTRL) algorithm is derived on the gradient descent approach. The PRWNN comprises a number of RWNN modules that are cascaded in a chained form and inherits the modular architectures of the pipelined recurrent neural network (PRNN) proposed by Haykin and Li. Since those modules of the PRWNN can be performed simultaneously in a pipelined parallelism fashion, it would result in a significant improvement in computational efficiency. And the performance of the PRWNN can be also further improved. Computer simulations have demonstrated that the PRWNN provides considerably better performance compared to the single RWNN model for nonlinear dynamic system identification.

Keywords Recurrent wavelet neural network · Pipelined recurrent neural network · Real-time recurrent learning · Nonlinear system identification

1 Introduction

Due to the nonlinear signal processing and learning capability by generating complex mapping between the input and the output space, artificial neural networks (ANNs) have become a powerful tool for nonlinear dynamic system identification [1]. Many research works using multilayer perceptron (MLP) networks [1], radial basis function (RBF) networks [2], functional link ANNs (FLANNs) [3],

H. Zhao (✉) · X. Zeng

School of Electrical Engineering, Southwest Jiaotong University, 610031 Chengdu, China
e-mail: hqzhao@home.swjtu.edu.cn

and recurrent neural networks (RNNs) [4] have been reported in nonlinear dynamic system identification.

Recently, the wavelet neural network (WNN), combining the capability of ANN in learning from processes and the capability of wavelet decomposition, has received considerable interest. In [5], a WNN based on the wavelet transform theory was presented as an alternative to ANNs for approximating nonlinear functions. Research results have shown that a WNN can approximate any continuous function over a compact set and have high accuracy and fast learning ability. However, it has been proved that the NN with the recurrent architecture is superior to feedforward neural network (FNN) in identifying nonlinear dynamic system. As a recurrent network, the recurrent wavelet neural network (RWNN) [6–8], combining the properties of attractor dynamics of the RNN and good convergence performance of the WNN, can cope with time-varying input or output through its own natural temporal operation because a mother wavelet layer composed of internal feedback neurons to capture the dynamic response of a system. To further improve the performance of the RWNN, the self-recurrent wavelet neural network (SRWNN) [9] and recurrent fuzzy wavelet neural network (RFWNN) [10, 11] have been presented to deal with the problems of nonlinear dynamic system identification. Although the RWNN shows promising results, it still suffers from the heavy computational loads as the RNN.

In 1995, to reduce the computational complexity of the RNN, a computationally efficient modular nonlinear adaptive filter-based pipelined recurrent neural network (PRNN) was proposed by Haykin and Li [12]. The design of the pipelined architecture follows the important engineering principle of divide and conquer and the biological principle of NN modules. Its significant merit is relatively low computational complexity [12–20]. As a result, inspired by the pipelined architecture of the PRNN, a novel modular RWNN based on the pipelined architecture is proposed to reduce the computational complexity and improve the performance of the RWNN for nonlinear dynamic system identification in this paper.

2 A Modular RWNN Based on the Pipelined Architecture

To overcome the computational complexity problem of the RWNN, keeping the views of the pipelined architecture, a novel modular RWNN based on the pipelined architecture (PRWNN) is presented. The PRWNN, inheriting the modular architectures of the PRNN proposed by Haykin and Li, comprises a number of RWNN modules that are cascaded in a chained form. Each module is implemented by a small-scale RWNN with internal dynamics. Since those modules of the PRWNN can be performed simultaneously in a pipelined parallelism fashion, it would result in a significant improvement in computational efficiency. In addition, the nesting module of the pipelined architecture can help to circumvent the problem of vanishing gradient of the RWNN, and the performance of the PRWNN can be further improved.

Figure 1 describes the structure of the PRWNN, which is composed of M identical modules, and each module is designed as a decision feedback RNN with q neurons and has $q - 1$ neuron output decision feedback to its input, and the remaining neuron output (the first neuron output decision) is applied directly to the next module. In the case of the PRWNN, module M is a fully connected RNN, and a one-unit delayed signal of the M module's output is assumed to be decision feedback to the input. Information flow into and out of the modules proceeds in a synchronized fashion. Therefore, all the modules have exactly the same number of external inputs and internal decision feedback signals.

Figure 2 shows the detailed structure of module i with q neurons and p external inputs. Note that for module M , its module output decision acts as an external feedback signal to itself. In addition, all the modules of PRWNN operate similarly in that they all have exactly the same number of external inputs and feedback signals, which are properly timed. Moreover, all the modules are designed to have exactly the same $(p + q + 1)$ -by- q synaptic weight matrix $W(n)$, q -by-1 weight vector $W^o(n)$ and the parameters of the wavelet function. An element $w_{k,l}(n)$ of this matrix represents the weight of the connection to the k th neuron from the l th input node. Moreover, the weight matrix W may be written as

$$W(n) = [w_1(n), \dots, w_k(n), \dots, w_q(n)]. \tag{1}$$

where $w_k(n)$ is a $(p + q + 1)$ -by-1 vector defined by

$$w_k(n) = [w_{1,k}(n), w_{2,k}(n), \dots, w_{p+q+1,k}(n)]^T. \tag{2}$$

And the superscript T denotes transposition.

At the n th time, for the i th module, the external input signal is described by the p -by-1 vector

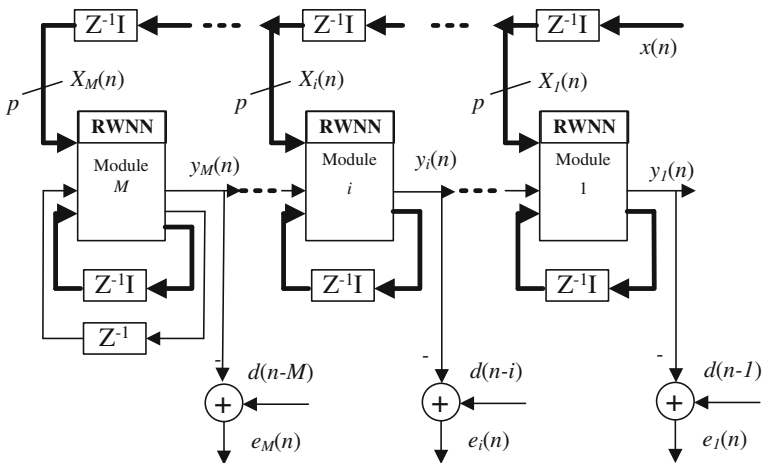


Fig. 1 A modular RWNN based on the pipelined architecture with M modules

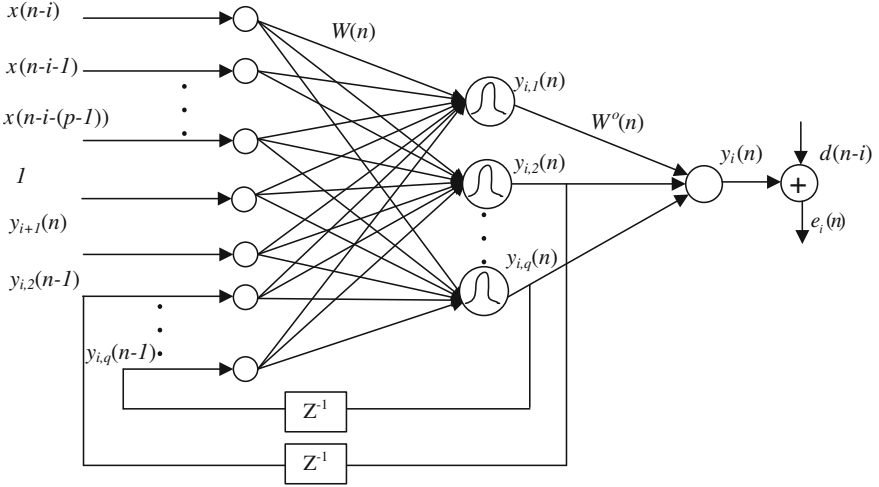


Fig. 2 Detailed architecture of module i of the PRWNN

$$X_i(n) = [x(n-i), x(n-(i+1)), \dots, x(n-(i+p-1))]^T. \quad (3)$$

and is delayed by $Z^{-i}I$ at the input of the module i , where Z^{-i} denotes the delay operator i time units, and I is the $(p \times p)$ -dimensional identity matrix, and p is the nonlinear adaptive equalizer order. The other input vector applied to module i is the q -by-1 decision feedback vector

$$r_i(n) = [y_{i+1,1}(n), \hat{r}_i(n)]^T, \quad i = 1, 2, \dots, (q-1). \quad (4)$$

where $y_{i+1,1}(n)$ is the first neuron's output in the adjacent module $i+1$, vector \hat{r}_i is the one-step delayed output feedback signals that originate from module i itself and is defined by

$$\hat{r}_i(n) = [y_{i,2}(n-1), \dots, y_{i,q}(n-1)]^T. \quad (5)$$

The last module of the PRWNN, namely module M , operates as a standard fully connected RWNN. The vector r_M consists of the one-step delayed output decision signals in module M that are fed back to itself and as shown by

$$\begin{aligned} r_M(n) &= [y_{M,1}(n-1), \hat{r}_M(n)]^T \\ &= [y_{M,1}(n-1), y_{M,2}(n-1), \dots, y_{M,q}(n-1)]^T. \end{aligned} \quad (6)$$

To accommodate a bias for each neuron, besides the $p+q$ inputs, the fixed input +1 is included. Based on the above discussion, an input vector $V_i(n)$ consisting of total $(p+q+1)$ input signals applied to module i is represented

$$V_i(n) = [X_i^T(n), 1, r_i^T(n)]^T, \quad i = 1, 2, \dots, M. \quad (7)$$

For the module i , the output $y_{i,l}(n)$ of neuron l at the n th time point is computed by passing $u_{i,l}(n)$ through a wavelet function $\varphi(\bullet)$, obtaining

$$y_{i,l}(n) = \varphi\left(\frac{u_{i,l}(n) - b_l(n)}{a_l(n)}\right). \quad (8)$$

With loss of generality, the ‘‘Gaussian-derivative’’ wavelet function given in [6] is used by

$$\varphi(x) = \frac{1}{\sqrt{|a_l|}} (-x) \exp\left(-\frac{x^2}{2}\right). \quad (9)$$

And the net internal activity $u_{i,l}(n)$ is given by

$$\begin{aligned} u_{i,l}(n) &= V_i^T(n)w_l(n) \\ &= \sum_{k=1}^{p+M+1} w_{k,l}(n)v_{i,k}(n) \\ &= \sum_{k=1}^p w_{k,l}(n)x(n - (i + k - 1)) \\ &\quad + w_{p+1,l}(n) + \sum_{k=p+2}^{p+q+1} w_{k,l}(n)r_{i,k-(p+1)}(n). \end{aligned} \quad (10)$$

where

$$v_{i,k}(n) = \begin{cases} x(n - (i + k - 1)), & 1 \leq k \leq p, 1 \leq i \leq M \\ 1, & k = p + 1, 1 \leq i \leq M \\ y_{i+1,1}(n), & k = p + 2, 1 \leq i \leq M - 1 \\ y_{M,1}(n - 1), & k = p + 2, i = M \\ y_{i,k-(p+1)}(n - 1), & p + 3 \leq k \leq p + 1 + q, 1 \leq i \leq M \end{cases} \quad (11)$$

and

$$r_{i,k-(p+1)}(n) = \begin{cases} y_{i+1,1}(n), & k = p + 2, 1 \leq i \leq M - 1 \\ y_{M,1}(n - 1), & k = p + 2, i = M \\ y_{i,k-(p+1)}(n - 1), & p + 3 \leq k \leq p + 1 + q, 1 \leq i \leq M \end{cases} \quad (12)$$

Then, the output of the i th module is given as

$$y_i(n) = \sum_{j=1}^q w_j^o(n)y_{i,j}(n) = H^T(n)W^o(n). \quad (13)$$

Finally, the output signal computed by the PRWNN at time instant n is defined by

$$y(n) = y_1(n). \quad (14)$$

Certainly, $y_i(n)$ is interpreted as the estimate of desired signal $d(n-i)$ computed by the i th module.

3 Training Algorithm for the PRWNN

According to the learning algorithms of the PRNN, adaptive learning algorithm of the PRWNN is derived by the real-time recurrent learning (RTRL) rule in the following subsection.

The overall cost function for the PRWNN is defined by

$$E(n) = \sum_{i=1}^M \varepsilon^{i-1} e_i^2(n). \quad (15)$$

where ε is an exponential forgetting factor that lies in the range of $0 < \varepsilon \leq 1$, the inverse of ε^{i-1} is a measure of the memory of the PRWNN. And the corresponding error $e_i(n)$ of the i th module is given by

$$e_i(n) = d(n-i) - y_i(n). \quad (16)$$

After every module of the PRWNN finishes its calculations, $e_1(n)$, $e_2(n)$, \dots and $e_M(n)$ error signals are obtained. Thus, adjustments to the synaptic weight matrix $W(n)$ and $W^o(n)$ of each module are made to minimize $E(n)$ in accordance with the RTRL algorithm.

According to the approach in [12], the change to k lth element of the weight matrix $W(n)$ is

$$\Delta w_{k,l}(n) = \frac{\eta_1}{2} \frac{\partial E(n)}{\partial w_{k,l}(n)}, \quad 1 \leq l \leq q, 1 \leq k \leq p+2+q. \quad (17)$$

Then, the element of weight matrix $W(n)$ is updated as

$$w_{k,l}(n+1) = w_{k,l}(n) + \Delta w_{k,l}(n) = w_{k,l}(n) - \frac{\eta_2}{2} \frac{\partial E(n)}{\partial w_{k,l}(n)}. \quad (18)$$

Similarly, the parameters $a_l(n)$ and $b_l(n+1)$ are updated by, respectively

$$a_l(n+1) = a_l(n) - \frac{\eta_3}{2} \frac{\partial E(n)}{\partial a_l(n)}. \quad (19)$$

$$b_l(n+1) = b_l(n) - \frac{\eta_4}{2} \frac{\partial E(n)}{\partial b_l(n)}. \quad (20)$$

Furthermore, according to the RTRL rule, the recursive equations of (18, 19) and (20) can be obtained by

$$w_{k,l}(n+1) = w_{k,l}(n) + \eta_2 \sum_{i=1}^M \varepsilon^{i-1} e_i(n) \left\{ \sum_{j=1}^q w_j^o(n) \frac{\varphi'(\text{net}_{i,j}(n))}{a_j(n)} \left[\frac{\partial y_{i+1,1}(n)}{\partial w_{k,l}(n)} w_{p+2,l}(n) \right. \right. \\ \left. \left. + \sum_{m=2}^q w_{m+p+1,j}(n) \frac{\partial y_{i,m}(n-1)}{\partial w_{k,l}(n)} + \delta_{k,j} v_{i,l}(n) \right] \right\}. \quad (21)$$

$$b_l(n+1) = b_l(n) - \eta_3 \sum_{i=1}^M \varepsilon^{i-1} e_i(n) \left\{ \sum_{j=1}^q w_j^o(n) \frac{\varphi'(\text{net}_{i,j}(n))}{a_j(n)} \left[w_{p+2,l}(n) \pi_- w_{k,l}^{i+1,1}(n) \right. \right. \\ \left. \left. + \sum_{m=2}^q w_{m+p+1,j}(n) \frac{\partial y_{i,m}(n-1)}{\partial w_{k,l}(n)} + \delta_{k,j} v_{i,l}(n) \right] \right\}. \quad (22)$$

$$a_l(n+1) = a_l(n) + \eta_4 \sum_{i=1}^M \varepsilon^{i-1} e_i(n) \left\{ \sum_{j=1}^q w_j^o(n) \varphi'(\text{net}_{i,j}(n)) \frac{\partial \left(\frac{u_{i,j}(n) - b_j(n)}{a_j(n)} \right)}{\partial a_l(n)} \right\}. \quad (23)$$

In addition, by using the gradient rule, the weight $W^o(n)$ of the PRWNN is updated as

$$W^o(n+1) = W^o(n) + \eta_1 \sum_{i=1}^M \varepsilon^{i-1} e_i(n) H_i(n). \quad (24)$$

where $\eta_i (i = 1, 2, 3, 4)$ is learning rate and controls the convergence performance of the PRWNN.

4 Simulations

To evaluate the performance of the PRWNN, nonlinear dynamic system identification application is carried out in this subsection.

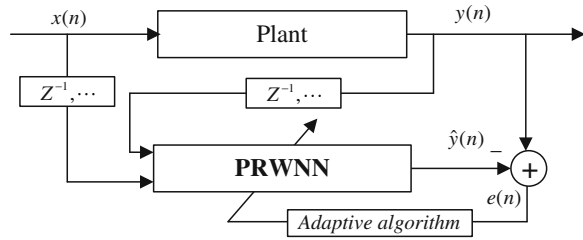
Figure 3 depicted the identification scheme of a nonlinear dynamic system based on the PRWNN filter. The plant is described by the following difference equation [6]

$$\hat{y}(n+1) = f[\hat{y}(n), \hat{y}(n-1), \hat{y}(n-2), x(n), x(n-1)]. \quad (25)$$

where the function $f(\cdot)$ is defined by

$$f[x_1, x_2, x_3, x_4, x_5] = \frac{x_1 x_2 x_3 x_5 (x_3 - 1) + x_4}{1 + x_3^2 + x_2^2}. \quad (26)$$

Fig. 3 Identification scheme of a nonlinear dynamic system



During the test phase, the following test signal is used to test the performance of the PRWNN models:

$$x(n) = \begin{cases} \sin(2\pi n/250) & 1 \leq n \leq 250 \\ 0.8 \sin(2\pi n/250) + 0.2 \sin(2\pi n/25) & < 250n \leq 600 \end{cases} \quad (27)$$

Fig. 4 Identification of the nonlinear dynamic plant with the test signal. **a** PRWNN. **b** RWNN

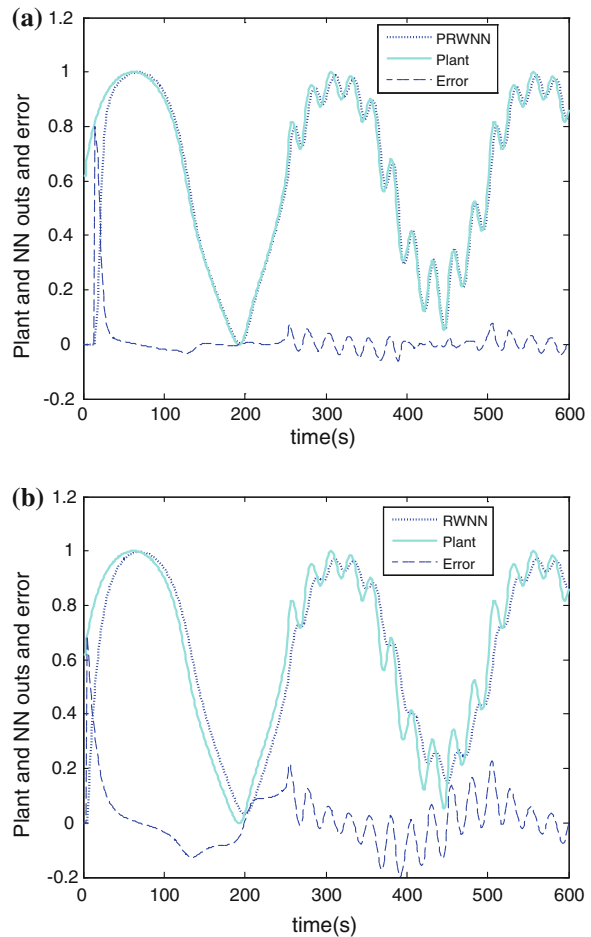


Figure 4 shows the actual neural network's output and error of the nonlinear dynamic plant for the test signal with the RWNN and PRWNN model. It is obviously observed that the proposed PRWNN shows much better performance than the conventional RWNN in this nonlinear dynamic system identification problem. This result is reasonable due to the fact that the pipelined architecture of the PRWNN helps to enhance nonlinear processing capability and improve the performance. Moreover, the computational complexity of the PRWNN is much lower than that of the RWNN.

5 Conclusion

In this paper, we proposed a nonlinear adaptive filter with a pipelined RWNN to reduce the computational burden of the RWNN. The network model consists of a number of modules-based RWNN that are interconnected in a chained form and inherits the major characteristics (low computational complexity) of the pipelined architecture. The parameter update rules of the PRWNN are derived according to the modified RTRL algorithm. The performance of the proposed PRWNN has been assessed for nonlinear dynamic system identification and compared with that of the RWNN model. Simulation results show that the proposed PRWNN with lower computational complexity can outperform the single RWNN model.

Acknowledgments This work was partially supported by the National Science Foundation of People's Republic of China (grant no: 61271340 and 61071183), Sichuan Provincial Youth Science and Technology Fund (grant no. 2012JQ0046), and Fundamental Research Funds for the Central Universities under grant SWJTU12CX026.

References

1. Narendra KS, Parthasarathy K (1990) Identification and control of dynamical systems using neural networks. *IEEE Trans Neural Networks* 1(1):4–27
2. Park J, Sandberg IW (1991) Universal approximation using radial basis function networks. *Neural Comput* 3:246–257
3. Patra JC, Pal RN, Chatterji BN, Panda G (1999) Identification of nonlinear dynamic systems using functional link artificial neural network. *IEEE Trans Syst Man Cybern B* 29(2):254–262
4. Wu YL, Song Q, Liu S (2008) A normalized adaptive training of recurrent neural networks with augmented error gradient. *IEEE Trans Neural Networks* 19(2):351–356
5. Zhang Q, Benviste A (1992) Wavelet networks. *IEEE Trans Neural Networks* 3(6):889–898
6. Rao SS, Kumthekar B (1993) Recurrent wavelet networks. In: *Proceedings of the 1993 IEEE-SP Workshop IEEE*, vol III, pp 3143–3147
7. Zhao F, Hu L, Li Z (2009) Nonlinear system identification based on recurrent wavelet neural network. In: *ISNN 2009*, vol 56, pp 517–525
8. Lu CH (2009) Design and application of stable predictive controller using recurrent wavelet neural networks. *IEEE Trans Ind Electron* 56(9):3733–3742

9. Yoo SJ, Choi YH, Park JB (2006) Generalized predictive control based on self-recurrent wavelet neural network for stable path tracking of mobile robots: adaptive learning rates approach. *IEEE Trans Circuits Syst I Regul Pap* 53(6):1381–1394
10. Abiyev RH, Kaynak O (2008) Fuzzy wavelet neural networks for identification and control of dynamic plants—a novel structure and a comparative study. *IEEE Trans Ind Electron* 55(8):3133–3140
11. Lin C, Chin C (2004) Prediction and identification using wavelet-based recurrent fuzzy neural networks. *IEEE Trans Syst Man Cybern B Cybern* 34(5):2144–2154
12. Haykin S, Li L (1995) Nonlinear adaptive prediction of nonstationary signals. *IEEE Trans Signal Process* 43(2):526–535
13. Mandic DP, Chambers JA (2000) On the choice of parameters of the cost function in nested modular RNNs. *IEEE Trans Neural Networks* 11(2):315–322
14. Zhao HQ, Zhang JS (2009) Nonlinear dynamic system identification using pipelined functional link artificial recurrent neural network. *Neurcomputing* 72:3046–3054
15. Stavrakoudis DG, Theocharis JB (2007) Pipelined recurrent fuzzy neural networks for nonlinear adaptive speech prediction. *IEEE Trans Syst Man Cybern B Cybern* 37(5):1305–1320
16. Zhao HQ, Zeng XQ, He ZY, Jin WD, Li TR (2012) Adaptive extended pipelined second-order Volterra filter for nonlinear active noise controller. *IEEE Trans Audio Speech Lang Process* 20(4):1394–1399
17. Zhao HQ, Zeng XQ, He ZY (2011) Low-complexity nonlinear adaptive filter based on a pipelined bilinear recurrent neural network. *IEEE Trans Neural Networks* 22(9):1494–1507
18. Zhao HQ, Zhang JS (2010) Pipelined Chebyshev functional link artificial recurrent neural network for nonlinear adaptive filter. *IEEE Trans Syst Man Cybern B Cybern* 40(1):162–172
19. Zhao HQ, Zeng XQ, Zhang JS, Li TR (2010) Nonlinear adaptive equalizer using a pipelined decision feedback recurrent neural network in communication systems. *IEEE Trans Commun* 58(8):2193–2198
20. Zhao HQ, Zhang JS (2009) A novel adaptive nonlinear filter-based pipelined feedforward second-order Volterra architecture. *IEEE Trans Signal Process* 57(1):237–246

Automated Text Data Extraction Based on Unsupervised Small Sample Learning

Yulong Liu, Shengsheng Shi, Chunfeng Yuan and Yihua Huang

Abstract Most of Web information extraction systems work with the DOM tree-based structured extraction rules to extract data records from Web pages; however, some, data items of, or even whole, of these data records are often in a semi-structured or unstructured text form. Thus, we need to introduce text data extraction rules to further extract the fine-grained data elements from those coarse-grained text items or records. However, generating text data extraction rules is a challenging task in either manual or automated way. In this paper, we propose an unsupervised learning approach to automatically deducing text data extraction rules from a small sample of text records. First of all, to prepare for extraction rule template deduction, we propose an iterative center core multiple sequence alignment method to align text columns in sample text records. Then, we propose an information entropy model based on the statistical features of text columns to further identify each column as either a template column or a data column. From identified template and data columns, plus some additional processing, we can quickly deduce the template, that is, the text data extraction rule. Eventually, we can use the text data extraction rule to perform the automated text data extraction from test text records. This unsupervised learning approach does not need any manual labeling and enables automated generation of text data extraction rules and text data extraction process. It is the first study effort toward the unsupervised small sample learning approach for automated text data extraction rule generation. The experimental results show that our approach achieves high accuracy.

Keywords Web information extraction · Data record · Template deduction · Small sample learning · Multiple sequence alignment · Text data extraction rule

Y. Liu · S. Shi · C. Yuan · Y. Huang (✉)
National Key Laboratory for Novel Software Technology,
Department of Computer Science and Technology, Nanjing University,
Nanjing 210093, China
e-mail: yhuang@nju.edu.cn

1 Introduction

The Web is the largest source of data information in the world. It contains a lot of useful and valuable information of interests to users or applications but unfortunately comes in unstructured and not well-organized form. Therefore, the problem of using effective means to extract information of interest from Web pages becomes an important research issue. In the last decade, many approaches have been reported for extracting information from Web pages [1–3]. Most of them work with the DOM tree-based structured extraction rules to extract data records from Web pages, such as Lixto [4], RoadRunner [5], MDR [6], and DEPTA [7]. But some, data items of, or even whole, of these data records are often in a coarse-grained and semi-structured or unstructured text form from which we need to further extract the fine-grained data elements.

Figure 1 shows a small sample of book records, each of which consists of a sequence of data elements. We need to extract data elements such as titles, authors, publication dates, and publishers from these flat text records. Therefore, it is necessary to introduce text data extraction rules to do this.

There are different approaches to generating text data extraction rules. One manual approach is to write rules similar to regular expressions by users. But this requires more professional skills for users. The second approach, such as Data-Mold [8], is the supervised learning by labeling a number of text instances and then training a model from the labeled instances. This approach is much better than

```

作者: (美) 克努特 (Knuth, D.E.) 著/2008-01-01/机械工业出版社
作者: (美) 克努特 著/2008-01-03/机械工业出版社
作者: (美) 克努特 (Knuth, D.E.) 著/2008-01-21/机械工业出版社
作者: (美) 高德纳 著/2010-10-31/人民邮电出版社
作者: (美) Donald E. Knuth 著/2002-09-28/清华大学出版社
作者: (美) 克努特 著, 苏运霖 译/2006-08-06/机械工业出版社
作者: (美) 高德纳 著/2010-10-01/人民邮电出版社
作者: (美) 克努特 著, 黄林鹏 译/2010-08-23/机械工业出版社
作者: (美) 克努特 编著, 苏运霖 译/2007-04-16/机械工业出版社
作者: 苏运霖/2002-09-19/国防工业出版社
作者: (德) 伯特特, (德) 卡斯特兰 著, 顾明 等译/2010-01-25/清华大学出版社
作者: 梁爽 等编著/2010-02-23/清华大学出版社
作者: (美) 威诺格拉德 (Winograd, T.) 等著, 韩柯 等译/2005-01-08/机械工业出版社
作者: (美) 皮特曼 (Pittman, T.), 皮特斯 (Peters, J.) 著/2010-01-21/机械工业出版社
作者: 范蓓 编著/2010-08-01/水利水电出版社
作者: 李丹 主编/2009-09-01/广东高等教育出版社
作者: 侯林 主编/2009-09-01/水利水电出版社
作者: 唐雯虹 编著/2008-03-01/清华大学出版社
作者: 李巍 著/2006-09-01/西南师范大学出版社
作者: 黄吉淳, 刘春 编著/2005-10-01/重庆大学出版社
作者: 章曙方 编/2006-01-01/上海人民美术出版社
作者: 黄元庆 编著/2007-05-01/东华大学出版社
作者: (美) 菲利奇 (Felici, J.) 著, 胡心怡, 朱琪颖 译/2006-01-01/上海人民美术出版社
作者: 徐航, 杨春晓 编著/2008-10-01/重庆大学出版社
作者: 肖清风, 黄淮 编/2007-09-01/重庆大学出版社
作者: 2005-08-01/中国建筑工业出版社
作者: (新) 莫里纽克斯 著, 李刚, 陈宇星 等译/2010-01-01/机械工业出版社

```

Fig. 1 A small sample of text records

the manual one but still brings some burden to users. The optimal approach is the unsupervised learning that allows the system to automatically deduce text data extraction rules from a small set of text instances. Unfortunately, to the best of our knowledge, few researches exist toward the unsupervised small sample learning approach for automated generation of text data extraction rules. Although there are some Web data extraction systems or research work like MDR [6], DEPTA [7], CTVS [9], and FiVaTech [10] that work on the basis of automated extraction rule deduction by unsupervised sample page learning and analysis, still they only deal with the DOM tree-based structured extraction rules without the ability to handle fine-grained text data extraction rules.

As shown in Fig. 1, usually text items in a set of sample text records contain certain pattern that we call a template. If we can deduce the template from the sample, we can use it to extract data elements from test text records. In our context, the deduced template is equal to a text data extraction rule.

To achieve the unsupervised rule generation, first we propose an *iterative center core* multiple sequence alignment method to align text items in the sample. The result output from the alignment process consists of a sequence of columns. Then, we adopt the concept of information entropy and propose an information entropy model to further identify each of the columns as either a template column or a data column. By applying a template deduction process on the identified columns, the text data extraction rule (template) will be generated. Eventually, we can use the automatically deduced rule to extract data elements from test text records.

The rest of the paper is organized as follows. Section 2 discusses related work. Section 3 describes the overview of our algorithm. Section 4 presents the iterative center core multiple sequence alignment method in detail for our algorithm. Section 5 discusses the template deduction process, generation of text data extraction rules and data extraction process. Section 6 presents our experiments. Section 7 concludes this paper.

2 Related Work

There are few researches on the unsupervised small sample learning approach for the generation of text data extraction rules. However, there are a few studies that are similar to our work on the basic problem. LISTEXTRACT [11] proposes an unsupervised approach based on statistical language models to extract relational tables from lists on the Web. It includes three major phases to perform whole processing: independently splitting text lines into fields, aligning the fields, and refining the alignment got from the last phase. To achieve automated processing, LISTEXTRACT requires two data sources: (1) a large-scale language model that records word co-occurrence scores and (2) a large corpus of automatically extracted HTML tables. The language model is used to identify candidate phrases that should not be split within a line, and the table corpus identifies phrases that occur elsewhere in table cells. Due to requirement of the large-scale language

model and corpus, obviously this approach is totally different from and not suitable for our approach.

The Phoebus [12] proposes a supervised method to extract text information from unstructured posts. Posts are the listings such as those found on eBay or Internet forum postings. The Phoebus system exploits reference sets to overcome the difficulty in lack of structure. A reference set is a relational dataset that contains entities and their attributes and is defined by users. The users also need to label the sample posts and examples of the extracted attributes. Thus, it is a semi-supervised method.

The named entity recognition is another research problem similar to our work. The main task of named entity recognition is to recognize all proprietary names and meaningful numbers that appear in the text and categorize them. The named entities include people's names, organization names, address, time expressions (date, time), and numeric expressions (currencies, percentage). Some approaches [8, 13, 14] to the named entity recognition are proposed based on HMM (hidden Markov model). The typical one is the system DataMold [8]. DataMold has a training phase that needs a fixed set of E elements in the form such as "House #," "Street," and "City" and a collection of T example text records that have been segmented into one or more of these elements. Thus, DataMold is a supervised system. Other named entity recognition methods or systems usually work in supervised way either. In addition, our task of extracting text elements from text records is different from a typical named entity recognition task. We need to identify the structure or pattern of a relatively short length of usually semi-structured text records from which we can extract all data elements, while the named entity recognition intends to identify and extract predefined data elements from usually long length of free text.

3 Major Process of our Algorithm

Automatically generating text data extraction rule is a process to deduce a template from a small sample of a number of text instances. For the sample in Fig. 1, human can easily recognize the template behind the sample, but it is not easy for programs to automatically recognize the template. To prepare for template deduction, first we propose an *iterative center core* multiple sequence alignment method to align the text items in the text record instances. As shown in Fig. 2, the alignment process will align all publication dates and author columns.

An efficient algorithm for the multiple sequence alignment we propose is iterative center core method. First, the center core method aligns original sample sequences into a set of aligned sequences. Then, the iterative step refines the alignment and gets optimized aligned sequences.

Based on the aligned text items, we will conduct the template deduction process. To recognize the template, we need to distinguish template columns from data columns. A template column consists of template elements, while a data

Fig. 2 A small sample of text records

作者	2008-01-01
作者	2008-01-03
作者	2008-01-21
作者	2010-10-31
作者	2002-09-28
作者	2006-08-06
作者	2010-10-01
作者	2010-08-23
作者	2007-04-16
作者	2002-09-19
作者	2010-01-25
作者	2010-02-23
作者	2005-01-08
作者	2010-01-21
作者	2010-08-01
作者	2009-09-01
作者	2009-09-01
作者	2008-03-01
作者	2006-09-01
作者	2005-10-01

column consists of data elements. A template element is the one that remains invariable or nearly invariable among all text records or text items, while a data element is the one that varies among all text records or text items. Obviously, the first column in Fig. 2 is a template column, and the second is a data column.

To distinguish template columns from data columns, we introduce the concept of information entropy and define an information entropy model based on the statistical features of these columns. The information entropy calculated for each column is used to determine whether a column is either a template column or a data column. We calculate information entropy for each column to measure the inconsistency of all elements that come from all record instances and belong to the column. The higher the information entropy is, the greater is the degree of inconsistency, and thus, the column is more likely to be a data column. Otherwise, the column is more likely to be a template column.

Last step with some postprocessing will eventually generate the template, that is, the text data extraction rule. When text data extraction rule is used to extract data elements from a test text record, we just match the test text record with the extraction rule. If an element in the test text is aligned with a template column, then the element is considered to be a template element. After identifying all the template elements, the remaining elements in the item are the data elements that we want to extract. The major process of our algorithm is shown in Fig. 3.

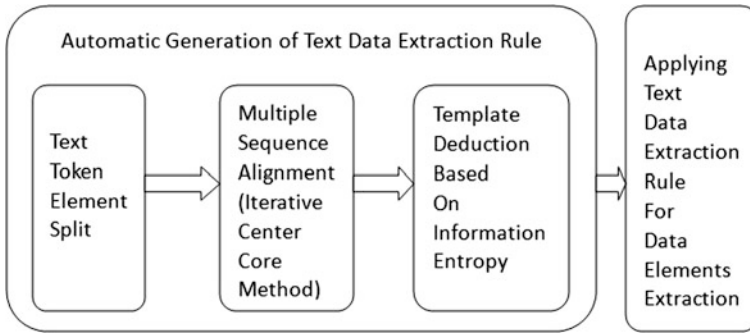


Fig. 3 Major process of our algorithm

4 Iterative Center Core Multiple Sequence Alignment Method

4.1 Text Token Element Split

The preprocessing we need to take to make preparation for multiple sequence alignment is to split text tokens for each of sample text records. In our algorithm, we split tokens with delimiters, and thus, tokens are delimiters and words in our algorithm. Figure 4 shows a sample text record and a sequence of split tokens. We call this sequence *original sequence*.

When splitting tokens, one thing needs to be pointed out is the recognition of entities. As shown in Fig. 4, “2010-11-8” is a date entity and we should not split it into five tokens of 3 numbers plus 2 “-”. To achieve this, we build a library of commonly used entities based on one or more regular expressions. These entities include number, date, time, currencies.

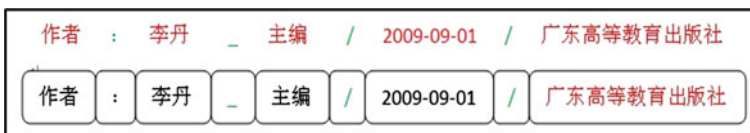


Fig. 4 A sample text record and its split tokens, that is, original sequence (here use “_” to represent a space)

4.2 Center Core Method

4.2.1 An Overview of the Method

After the token split preprocessing, all sample text records are converted into a set of original sequences, $\{S_i, i = 1, 2, \dots, k\}$. Then, inspired by the center star method [15], we propose the center core method to conduct the multiple sequence alignment processing.

Basically, there are two rounds of alignment processes. In the first round, the center core method first selects a *core sequence* from the set of original sequences. To achieve this, we make alignment for every pair of original sequences in the set and obtain a similarity score $f(s_i, s_j)$. Then, for each original sequence s_i , there is an accumulated similarity score $\sum_{j \neq i} f(s_i, s_j)$, where $j = 1, 2, \dots, k$ and $j \neq i$. Then, we sort the original sequences by their accumulated similarity scores from high to low and take the first original sequence as the core sequence.

In the second-round alignment, we align a number of original sequences together to generate a set of *aligned sequences* that will be used later for template deduction processing.

4.2.2 Score Function for Core Sequence Selection

In the first-round alignment, we make alignment between every two of original sequences to select core sequence. To do this, we need to define the score functions to measure the similarity between tokens from two original sequences and the similarity between two original sequences.

First, the similarity between different types of tokens is defined as 0. The similarity between two different delimiters is defined as 0 as well. For example, “,” and “.” have similarity 0. To determine the similarity between two words, we need to make an alignment between these two words (treat them as strings) and get a score. The score function for string alignment is described as the following formula:

$$\text{gain}(c_i, c_j) = \begin{cases} 1, & c_i, c_j \text{ are the same characters;} \\ 0.5, & c_i, c_j \text{ are of the same type;} \\ 0, & \text{otherwise;} \end{cases} \quad (1)$$

For example, the alignment between “等编著” and “编著” has a score 2. Then, for normalization, we divide the score with the length of the longer word and obtained the similarity $2/3$. Instead of simply determining the similarity between two same delimiters, we need to consider their context. A word before a delimiter will be treated as part of the delimiter’s context. If two delimiters are the same, then the similarity between these two delimiters’ contexts is used as the similarity

between these two same delimiters. For example, there are two “,” and two words before them are “著” and “编著”. The similarity between the two “,” is 0.5, which is the similarity between “著” and “编著”.

After giving the similarity function between tokens, we define the score function for sequence alignment as follows:

$$\text{gain}(e_i, e_j) = \begin{cases} \text{similarity}(e_i, e_j) \times 100, & e_i, e_j \text{ are two words;} \\ \text{similarity}(e_i, e_j) \times 20, & e_i, e_j \text{ are two identical} \\ & \text{delimiters;} \\ 0, & e_i \text{ or } e_j \text{ is a space;} \\ -100, & \text{otherwise;} \end{cases} \quad (2)$$

where e_i and e_j are token elements, either a word or a delimiter. Our score function gives a penalty of -100 for the alignment of them so that we will not get a column that contains both delimiters and words.

As described in the overview of the method in this section, given the score function, we will make the alignment process to select the core sequence. Intuitively, the core sequence is the original sequence that is most similar to all rest of the other original sequences in the set.

4.2.3 Multiple Sequence Alignment

After obtaining the core sequence, we need to perform a multiple sequence alignment process to align all original sequences together to generate a set of aligned sequences so that we can deduce the template from the aligned sequences in the next step. The specific process for the multiple sequence alignment is as follows:

- (1) Put the core sequence into the aligned sequences (the initial set of aligned sequences is empty, and thus, now the set of aligned sequences contains only one original sequence).
- (2) Pick one original sequence from the remaining original sequences and align it to the aligned sequences.
- (3) If the set of original sequences is not empty go to step (2), otherwise stop the process.

Figure 5 shows a set of aligned sequences resulted from the second-round alignment.

In step (2), we need to align and add a token element to a column. Now, suppose there is a column C in the aligned sequences and a token element e belonging to an original sequence that needs to be aligned and added to the aligned sequences, where C contains a set of aligned token elements $e_1, e_2, e_3, \dots, e_w$ that come from each of the aligned original sequences. To align and add the original sequence to the aligned sequences, we align C and e together:

$$e_1, e_2, e_3, \dots, e_w, e.$$

作者	:	范蓓	-	编著	/	2010-08-01	/	水利水电出版社
作者	:	李丹	-	主编	/	2009-09-01	/	广东高等教育出版社
作者	:	侯林	-	主编	/	2009-09-01	/	水利水电出版社
作者	:	唐虹	-	编著	/	2008-03-01	/	清华大学出版社
作者	:	李巍	-	著	/	2006-09-01	/	西南师范大学出版社

Fig. 5 An example of a set of aligned sequences and the first line was selected as the core sequence

Then, the score function for this alignment is as follows:

$$\text{gain}(C, e) = \sum_{1 \leq i < j \leq w} \text{gain}(e_i, e_j) + \sum_{i=1}^w \text{gain}(e_i, e) \tag{3}$$

In actual programming, $\sum_{1 \leq i < j \leq w} \text{gain}(e_i, e_j)$ is an attribute of C . When calculating $\text{gain}(C, e)$, we can directly access $\sum_{1 \leq i < j \leq w} \text{gain}(e_i, e_j)$. Thus, only $\sum_{i=1}^w \text{gain}(e_i, e)$ needs to be calculated.

4.3 Iterative Center Core Method

By applying the center core method, we get a set of aligned sequences S . As mentioned before, every column C in the aligned sequences has an attribute $\sum_{1 \leq i < j \leq w} \text{gain}(e_i, e_j)$. Accumulating the attributes from all columns gives an overall score $\text{score}(s)$ for S . This overall score reflects how well the original sequences are aligned together into a set of aligned sequences. However, the effect of alignment is impacted by the order that each of original sequences is added to the aligned sequences. To reduce the impact and further improve the accuracy of alignment, we propose an iterative center core method for further optimization of the center core alignment method.

Suppose that the aligned sequences S consist of $w + 1$ original sequences, s_1, s_2, s_w, s_{w+1} , where s_1 is the core sequence. The iterative center core method does the following in each iteration:

Pick s_1 from S , and now, S consists of w original sequences. Then, align s_1 with S . After this step, S is in the form of $s_2, \dots, s_w, s_{w+1}, s_1$. Repeating this with $w + 1$ steps, we complete one iteration and get new aligned sequences s^2 .

It should be noted that by $w + 1$ rotation, original sequences in s^2 are still in the order of $s_1, s_2, \dots, s_w, s_{w+1}$. By continuing the iteration, we get a sequence of $s^2, s^3, s^4, \dots, s^n, s^{n+1}$. The sequence $\text{score}(s), \text{score}(s^2), \dots, \text{score}(s^n), \text{score}(s^{n+1})$ is monotonically increasing and convergent. In another word, our iterative method will terminate in the best result. Here, we will briefly illustrate the above conclusion. As mentioned before, in each iteration of the algorithm, we do $w + 1$ steps. (the aligned sequences S consist of $w + 1$ sequences). After every step, we

get aligned sequences S' from S . And it can be easily proved that $\text{score}(s') > \text{score}(s)$. As a corollary, we get $\text{score}(s^{i+1}) > \text{score}(s^i)$. Thus, the sequence is monotonically increasing.

The pseudo-code of the iterative center core method is as follows:

```

Iterative Center Core(AlignedSequences S)
// Input: A set of aligned sequences S
// Output: Aligned sequences after iterative optimization
height=S.getHeight()
do
  oldScore=Score(S)
  for i=1 to height do
    remove the first original sequence s from S
    align(S,s)
  endfor
  newScore=Score(S)
while newScore>oldScore
return S

Score(AlignedSequences S)
// Input: A set of aligned sequences S
// Output: Score for this alignment
Score=0
for each column C in S do
  Score=Score+Score(C)
endfor
return Score

Score(Column C)
// Input: A column C
// Output: Score for the alignment of elements in C
score=0
for each element  $e_i$  in C do
  for each element  $e_j$  in C do
    if  $i < j$  do
      score=score+gain( $e_i, e_j$ )
    endif
  endfor
endfor
return score

```

5 Template Deduction and Text Data Extraction

5.1 Identify Columns Based on Information Entropy

After completing the multiple sequence alignment, we get a set of aligned sequences. Now, before starting to deduce template, we need to identify each column in the aligned sequences as either a template column or a data column.

Figure 2 shows a template column and a data column. We can observe that token elements in a template column tend to be constant or nearly constant, while token elements in a data column vary dramatically. In other word, the inconsistency of token elements belonging to a data column is high while that of token elements belonging to a template column is relatively low. Thus, we can distinguish template columns from data columns, depending on the inconsistency of token elements. To measure the inconsistency of token elements belonging to a column, we introduce the conception of information entropy [16]:

$$\begin{aligned} H(x) &= E[\log(1/p(x_i))] \\ &= - \sum p(x_i) \log(p(x_i)) \quad (i = 1, 2, \dots, n) \end{aligned} \quad (4)$$

$H(X)$ denotes the entropy of a discrete random variable X with possible values $\{x_1, \dots, x_n\}$ and probability mass function $p(X)$ ($p(x_i) = \Pr(X = x_i), i = 1, 2, \dots, n$). Information entropy is a measure of the uncertainty associated with a random variable.

Now in our case, the random variable X is the token element appearing in a column. And a token element has many values. $p(x_i)$ becomes the ratio of a token element x_i to all token elements of the column.

To calculate $p(x_i)$, we need to partition all token elements into equivalence classes. An equivalence class is a set that consists of equivalent token elements. Assume that all token elements are partitioned into n equivalence classes: E_1, \dots, E_n . We call the number of token elements contained in an equivalence class as the modulus of the equivalence class, denoted by $|E|$. Then, $p(x_i)$ will be calculated as: $p(x_i) = \frac{|E_i|}{\text{sum}}, x_i \in E_i$. Further, the information entropy of the token elements in a column C will be calculated as follows:

$$H(C) = - \sum \frac{|E_i|}{\text{sum}} \log\left(\frac{|E_i|}{\text{sum}}\right), \text{sum} = \sum |E_i| \quad (5)$$

If the inconsistency of token elements of a column is high, the calculated information entropy will be high. Otherwise, the information entropy will be low. Because the inconsistency of a data column is much higher than that of a template column, this will allow us to determine whether a column is a template column or a data column in terms of its calculated information entropy. We set a threshold for the determination.

E_{n+1} }. Then, the information entropy of S' will be $H(C_{S'})$. It is easy to see that $H(C_{S'}) < H(C_S)$ since $-\frac{|E_{n+1}|}{\text{sum}} \log\left(\frac{|E_{n+1}|}{\text{sum}}\right) < -\frac{|E_i|}{\text{sum}} \log\left(\frac{|E_i|}{\text{sum}}\right) - \frac{|E_j|}{\text{sum}} \log\left(\frac{|E_j|}{\text{sum}}\right)$ here, $|E_{n+1}| = |E_i| + |E_j|$

Thus, the information entropy of the column in Fig. 6 is lowered after the above process. As a result, the column is identified as a template column. By the way, although the above process tends to lower the information entropy of a column, the calculation of information entropy for a data column will hardly be affected. In a data column, the case that a token element is a substring of another token element rarely happens. Thus, the information entropy of a data column will not be lowered or be lowered a little by the above process.

The calculation algorithm for the information entropy of a column is listed as follows:

```

Entropy(Column C)
// Input: A column
// Output: The column's information entropy
entropy=0
height=C.getHeight()
Set S={}
Set Q={}
for each element e in C do
    if e is not a sapce
        copy=e.clone()
        add copy to S
    endif
endfor
sum=S.size()
existRate=sum/height
if existRate<existRateThreshold
    entropy=largeEntropy
    return entropy
Endif
Q=PartitionIntoEquivalenceClasses(S)
while Q is not empty do
    remove some equivalence class E from Q

    entropy = entropy -  $\frac{|E|}{\text{sum}} \log\left(\frac{|E|}{\text{sum}}\right)$ 
endwhile
return entropy

```

5.2 Template Deduction

After calculating the information entropy for each column and determining template columns and data columns, plus some subsequent processing, we will obtain the deduced template. Figure 7 shows an example of deduced template from the sample example in Fig. 1. The deduced template is “作者:XXX 著,XXX 等译/

XXX/XXX,” in which XXX indicates a data column that we need to extract later. The rest of the strings in this template will work as template columns that will guide the matching process when applying this template as the extraction rule to extract data elements from test text records.

5.3 Data Element Extraction

The deduced template will be used as the text data extraction rule. After obtaining the text data extraction rules, we can apply it to extract data elements from test text records. The basic process for this is still sequence alignment. We will split a test text record with delimiters into an original sequence and then align it with template columns in the text data extraction rule. If a token element matches with a template column, then the element is considered as a template element. After identifying all the template elements, the remaining elements in the sequence are the data elements that we want to extract. In Fig. 7, “作者”, “:”, “_”, “著”, “/”, and “/” are identified as template elements but “(新)莫里纽克斯”、“李刚,陈宇星”, “2010-01-01”, and “机械工业出版社” are identified as data elements.

6 Experiments

We carried out experimental studies to evaluate the accuracy of our proposed method.

Datasets: We obtain 10 datasets from 10 Websites, with each dataset containing 100 different text records. From each dataset, we take 30 text records as samples to deduce the text data extraction rule and then apply the deduced rule to extract data elements from the dataset.

Evaluation: For each dataset, we first manually identify the data elements and count the total number of data elements belonging to the dataset. After extraction, we count the total number of extracted data elements and the number of correctly extracted data elements. We denote the total number of data elements belonging to a dataset as D_g . D is the total number of data elements extracted from the dataset. D_{correct} denotes the total number of correctly extracted data elements from the dataset.

Precision (P), recall (R), and f-measure (F) are calculated, respectively, as follows:

$$P = \frac{D_{\text{correct}}}{D} \quad R = \frac{D_{\text{correct}}}{D_g} \quad F = \frac{2 \times P \times R}{P + R}$$

Experimental results of applying generated rules to extract text data from the datasets are listed in Table 1. Further the accuracy evaluation is listed in Table 2.

作者: XXX	_著	/	XXX	/	XXX
作者: XXX	_著	/	XXX	/	XXX
作者: XXX	_著	/	XXX	/	XXX
作者: XXX	_著	/	XXX	/	XXX
作者: XXX	_著	/	XXX	/	XXX
作者: XXX	_著	, XXX	_译	/	XXX / XXX
作者: XXX	_著	, XXX	_译	/	XXX / XXX
作者: XXX	_著	, XXX	_译	/	XXX / XXX
作者: XXX	_编著	, XXX	_译	/	XXX / XXX
作者: XXX	_	/	XXX	/	XXX
作者: XXX	_著	, XXX	_等译/	XXX	/ XXX
作者: XXX	_等编著	/	XXX	/	XXX
作者: XXX	_等著	, XXX	_等译/	XXX	/ XXX
作者: XXX	_著	/	XXX	/	XXX
作者: XXX	_编著	/	XXX	/	XXX
作者: XXX	_主编	/	XXX	/	XXX
作者: XXX	_主编	/	XXX	/	XXX
作者: XXX	_编著	/	XXX	/	XXX
作者: XXX	_著	/	XXX	/	XXX
作者: XXX	_编著	/	XXX	/	XXX
作者: XXX	_编	/	XXX	/	XXX
作者: XXX	_编著	/	XXX	/	XXX
作者: XXX	_著	, XXX	_译	/	XXX / XXX
作者: XXX	_编著	/	XXX	/	XXX
作者: XXX	_编	/	XXX	/	XXX
作者:			XXX	/	XXX
作者: XXX	_著	, XXX	_等译/	XXX	/ XXX
作者: (新) 莫里纽克斯	_著	, 李刚, 陈宇星	_等译/	2010-01-01/	机械工业出版社

Fig. 7 Deduced template from the example in Fig. 1

Table 1 Results of extraction

Dataset	Dataset source website	D	Dcorrect	Dg
1	dangdang.com	335	335	341
2	china-pub.com	440	440	545
3	pconline.com.cn	1020	790	870
4	verycd.com/base/movie	515	515	515
5	e.360buy.com	152	145	245
6	ceppbooks.sgcc.com.cn	353	325	438
7	sciencedirect.com	450	450	450
8	verycd.com/base/game/	300	300	300
9	alibaba.com	538	403	485
10	springerlink.com	206	206	206
Sum		4309	3,909	4,395

Table 2 Statistical results of experiment

Dataset	Precision	Recall	f-measure
1	1.00	0.98	0.99
2	1.00	0.81	0.90
3	0.77	0.91	0.83
4	1.00	1.00	1.00
5	0.95	0.59	0.73
6	0.92	0.74	0.82
7	1.00	1.00	1.00
8	1.00	1.00	1.00
9	0.75	0.83	0.79
10	1.00	1.00	1.00
Sum	0.91	0.89	0.90

Experimental results analysis: Generally, the experimental results show that the unsupervised small sample learning approach for automated generation of text data extraction rules we proposed in this paper can reach high accuracy for most of the cases. Datasets 4, 8, and 10 are collections of relatively well-structured text items, and thus, the precision and recall for extraction on them reach 100 %.

However, the accuracy of our algorithm on dataset 5 is relatively low. The sample in dataset 5 contains 30 text items. And part of their aligned sequences is shown in Fig. 8.

Fig. 8 Part of text data records. From dataset 5

世阔	著 / 2002-01-01 / EPUB 格式
张登魁	著 / 2002-04-10 / EPUB 格式
冷玉兰	著 / 2009-04-01 / EPUB 格式
刘玉栋	著 / 2005-12-01 / EPUB 格式
东方闻睿	著 / 2004-05-04 / EPUB 格式
	EPUB 格式
山立, 等等	著 / EPUB 格式
孙新峰	著 / EPUB 格式
张文翰	著 / 2009-06-01 / EPUB 格式
李妍	著 / 2010-02-01 / EPUB 格式
	2008-05-01 / EPUB 格式
	2008-05-01 / EPUB 格式
林语堂	著 / 2005-10-01 / EPUB 格式
	2008-04-01 / EPUB 格式
	2008-05-01 / EPUB 格式
	2008-01-01 / EPUB 格式
侣海岩	著 / EPUB 格式
	2009-12-01 / PDF 格式
牛翁	著 / 2010-12-01 / PDF 格式
(宋)苏轼	著 / EPUB 格式
邹容	著 / EPUB 格式
	2012-07-12 / PDF 格式

The last column is originally a data column. However, there is a special feature of this column that it only contains two types of data elements: EPUB and PDF. This situation violates the observation in Sect. 5 that the token elements belonging to a data column vary dramatically. Thus, we get a low information entropy value, which leads to incorrect identification of the column and further leads to a low recall for final data extraction.

7 Conclusion

There are few research efforts toward the unsupervised small sample learning approach for automated generation of text data extraction rules, and our work is the first study effort on this problem. In this paper, we propose an unsupervised learning approach to automatically deduce text data extraction rules from a small sample and then apply the text data extraction rules to extract data elements from text records. Generating text data extraction rules from a small sample is equivalent to the deduction of template behind the sample. To deduce the template, first we propose a two-round center core multiple sequence alignment method to align all the original sequences of the sample. To further optimize the alignment method to achieve higher accuracy, we further propose an iterative center core multiple sequence alignment method to generate aligned sequences. Then, in order to determine template columns and data columns for template deduction, we introduce and define the information entropy based on the statistical features of text elements in the column of the aligned sequences. Finally, we can deduce templates that can be used as the text data extraction rules to conduct the data element extraction from test text records. The experimental results demonstrate that our proposed method and algorithm can provide high accuracy to automatically extract text data elements. Our approach is automated with no user intervention when generating text data extraction rules. It can work as a complement with those Web data extraction systems with the DOM tree-based structured extraction rules to perform fine-grained text data extraction.

Acknowledgments This work is funded by China NSF Grant (#61072152) and Jiangsu Province Industry Promotion Program (#BE2011172).

References

1. Laender AH, Ribeiro-Neto BA, da Silva AS, Teixeira JS (2002) A brief survey of web data extraction tools. *SIGMOD* 31(2):84–93
2. Boronat XA (2008) A comparison of HTML-aware tools for Web Data extraction. Leipzig
3. Kuhlins S, Tredwell R (2002) Toolkits for Generating Wrappers. *NetObjectDays*, 184–198
4. Baumgartner R, Gatterbauer W, Gottlob G (2001) Web data extraction system with Lixto. VLDB

5. Crescenzi V, Mecca G, Merialdo P (2001) RoadRunner: towards automatic data extraction from large web sites. VLDB, 109–118
6. Liu B, Grossman RL, Zhai Y (2003) Mining data items in Web pages. KDD, 601–606
7. Zhai Y, Liu B (2005) Web data extraction based on partial tree alignment. WWW. 76–85
8. Borkar V, Deshmukh K, Sarawagi S (2001) Automatic segmentation of text into structured records. SIGMOD 30(2):175–186
9. Su W, Wang J, Lochovsky FH, Liu Y (2011) Combining tag and value similarity for data extraction and alignment. TKDE 24(7):1186–1200
10. Kaye M, Chang C-H (2010) FiVaTech: page-level web data extraction from template pages. TKDE 22(2):249–263
11. Elmeleegy H, Madhavan J, Halevy A (2009): Harvesting relational tables from lists on the web. In: VLDB endowment. 2:1, pp 1078–1089
12. Carrillo H, Lipman D (1988) The multiple sequence alignment problem in biology. SIAM J Appl Math 48:1073–1082
13. Sun J, Zhou M, Gao J (2003) A class-based language model approach to chinese named entity identification. In: The association for computational linguistics and Chinese language processing, pp 1–28
14. Chua T-S, Liu J (2002) Learning pattern rules for Chinese named entity extraction. AAAI
15. Gusfield D (1993) Efficient methods for multiple sequence alignment with guaranteed error bounds. Bull Math Biol 55(1):141–154
16. Shannon CE (2001) A mathematical theory of communication. In: Mobile computing and communications review, pp 3–55

SLAM and Navigation of a Mobile Robot for Indoor Environments

Shuangshuang Lei and Zhijun Li

Abstract In this paper, a real-time implementation of simultaneous localization and mapping (SLAM) and navigation is described based on a mobile service robot platform, which consists of two driving wheels, a laser, and a Kinect. The main algorithm is extended Kalman filter (EKF) which is combined with feature extraction from laser scan data and extended beam curvature method for obstacle avoidance.

Keywords SLAM · Kalman filter · Obstacle avoidance

1 Introduction

Mobile service robots are very promising application of robotics. They should help people to deal with various tasks. The autonomous movement is the primary capability, so to let robot know where it is and how to achieve the goal would be fundamental problems of the mobile robot. To solve this problem, there are several simultaneous localization and mapping (SLAM) approaches, such as Kalman filter-based algorithm [1], Monte Carlo localization algorithm [2].

Monte Carlo localization algorithm is a probabilistic method which utilizes a samples set (particles) to approximate the probability density functions from a Bayesian perspective. This method had successfully introduced several indoor navigation applications. But the disadvantage is the high computational requirements.

Kalman filters are a efficient algorithm and widely introduced to solve the SLAM [3, 4]. However, the regular Kalman filter only can cope with linear

S. Lei · Z. Li (✉)

College of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, Guangdong, People's Republic of China

e-mail: zjli@ieee.org

problem. In Smith, Self, Cheeseman paper [5], they proposed the extended Kalman filter(EKF) to incrementally estimate the posterior distribution over robot pose along with the positions of the landmarks, through a Taylor expansion to linearize nonlinear system.

Obstacle avoidance is an another importance issue studied by researchers. Artificial potential field method [6], the neural network method [7], vector field histogram method (VFH) [8], the curvature velocity method (CVM) [9], the lane curvature method (LCM) [10], and the beam curvature method (BCM) [11] are typical techniques for obstacle avoidance. Among these techniques, beam curvature method is a simple and efficient method to find the promising forward path, which combined the LCM with CVM.

In this paper, we combine several methods to implement our algorithm. First, we utilize split–merge method to extract the essential features as the landmarks. Then, a EKF is used to solve the SLAM problem. Moreover, a modified BCM is adopted to avoid obstacles. At last, a kinematic control law is used to command the motion of the robot.

2 The Proposed Robot System

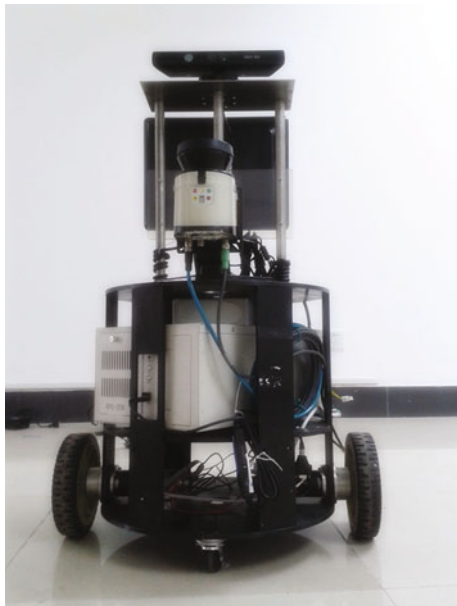
2.1 Mobile Service Robot

The main task of proposed system is to help operators to collect external environment information and build environment map in real time without collisions in the whole process. Our developed mobile robot platform is illustrated in Fig. 1. It equips two sensors, SICK LMS111 Laser Finder and Kinect, to grasp external information. The two wheels are driven by servo motors, and each servo motor is controlled by an ELMO driver which communicates with host through Kvaser CAN device.

3 Simultaneous Localization and Mapping

SLAM is a process which estimates the robot's pose and landmark's position at the same time, namely location and mapping. The fundamental step is to obtain external information to search proper features (landmarks). After landmarks' extraction, it is essential to match the observed landmarks with the existed landmarks in map. The heart of our SLAM process is correlating the mean and covariance matrix of system state through EKF technique.

Fig. 1 The developed mobile service robot



3.1 Feature Extraction

For the indoor environment, there exist many line features; accordingly, we choose the corners as landmarks. There are several common methods to extract line features [12]. In this paper, we use split–merge method [13] to extract lines and compute the corners' position.

The basic idea of split–merge is to repeatedly search a point with the maximum distance d_p to a line, which is decided by the start and end points of a dot set, and compare d_p with the threshold $d_{\text{threshold}}$. If d_p is lesser than $d_{\text{threshold}}$, the point P is considered as the point in the line, otherwise divide the dot set into two new sets at point P . After split, we utilize least square method to fit corresponding lines.

We can consider the fitting problem as a regression problem under polar coordinate using the raw laser data. So we can describe line as

$$L : \rho = x \cos(\alpha) + y \sin(\alpha), \quad (1)$$

where ρ is the perpendicular distance from the origin to the line and α is the angle between the x axis and the normal of the line. So the mean-squared error of distance is obtained as follows:

$$\sum_{i=1}^N d_i^2 = \sum_{i=1}^N (\rho - (x_i \cos(\alpha) + y_i \sin(\alpha)))^2, \quad (2)$$

where d_i is the distance from point (x_i, y_i) to the line L . x_i and y_i , respectively, are the positions of samples in Cartesian coordinates. α and ρ can be computed as follows [14]

$$\tan(2\alpha) = \frac{2 \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^N [(x_i - \bar{x})^2 - (y_i - \bar{y})^2]}, \quad (3)$$

$$\rho = \bar{x} \cos \alpha + \bar{y} \sin \alpha, \quad (4)$$

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i, \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i. \quad (5)$$

Combining with these two methods, we obtain the extracted line features. Through computing the cross dot of every two lines and comparing its position with the laser data, we can decide which cross dot could be corner. Moreover, we make some constraints, for example, a line should not be extracted if its length is under 1 meter or the number of contained dot is lesser than 10. The extracted results are shown in Figs. 2 and 3.

In Fig. 2, the blue dots are the raw laser data, the red circle represents robot, the red lines are extracted lines, and the green dot is the computed corner.

3.2 Data Association

In SLAM applications, when the robot detects a feature, it must be associated with a landmark in the existed map or incorporated as a new landmark. Data association

Fig. 2 The result without constraints

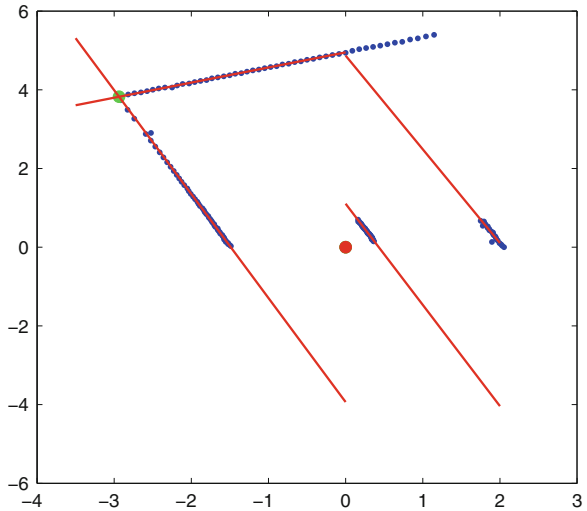
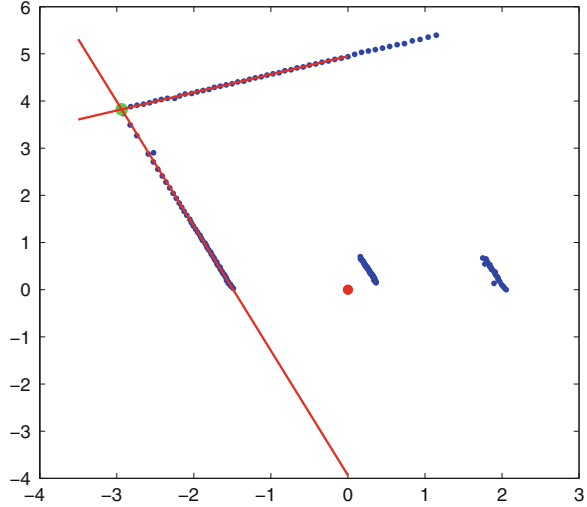


Fig. 3 The result with constraints



is the process that deals with the matching problem between observations and existed landmarks. A typical data association algorithm is composed of two elements: compatibility test to find potential pairs between observation and landmarks and a selection rule to choose the best matchings among the set of compatible matchings.

Generally, the gated nearest neighbor (NN) algorithm is used to associate data. The main process of NN algorithm is compatibility test through utilizing the Mahalanobis distance to determine whether a feature corresponds to a landmarks and then find the one with the minimum Mahalanobis distance. However, in this paper, due to that the landmarks are set exactly, we use a simple method to determine the correspondence between features and landmarks. We compute the degree of matching through a simple computation as follows:

$$\text{Match}(i, j) = |\text{error}_r| \cos(\text{error}_b), \tag{6}$$

where error_r and error_b are the error of range and bearing between feature i and landmark j , respectively. Then, we regard the feature with the minimum match as the best match of corresponding landmark.

3.3 Extended Kalman Filter

Extended Kalman filter is a repetitive process of prediction and correction to estimate the system state. In our implementation, we can estimate the pose of robot by observing several set landmarks.

$$x_n(k) = [x(k), y(k), \theta(k)]^T \tag{7}$$

where $x_n(k)$ is the posture of robot, $x(k)$ and $y(k)$ are the current robot positions, $\theta(k)$ is the heading angle, and k is the current time. The developed robot platform is a nonholonomic two-wheeled driven mobile robot, and its velocity motion model is shown in Fig. 4. In Fig. 4, D is the distance between two wheels. So velocity motion model can be described as

$$g(x, u, k + 1) = \begin{bmatrix} x(k) + \Delta t V(k) \cos(\theta(k)) \\ y(k) + \Delta t V(k) \sin(\theta(k)) \\ \theta(k) + \Delta t \omega(k) \end{bmatrix} \tag{8}$$

where $V(k)$ and $\omega(k)$ are the linear velocity and angular velocity, respectively.

To fit EKF, we should linearize the motion model to acquire the Jacobian matrices $\frac{dg}{dx}$ and $\frac{dg}{du}$ as follows:

$$G_x = \frac{dg}{dx} = \begin{bmatrix} 1 & 0 & -\Delta t V(k) \sin(\theta(k)) \\ 0 & 1 & \Delta t V(k) \cos(\theta(k)) \\ 0 & 0 & 1 \end{bmatrix} \tag{9}$$

$$G_u = \frac{dg}{du} = \begin{bmatrix} \frac{R}{2} \Delta t \cos(\theta(k)) & \frac{R}{2} \Delta t \cos(\theta(k)) \\ \frac{R}{2} \Delta t \sin(\theta(k)) & \frac{R}{2} \Delta t \sin(\theta(k)) \\ \frac{R}{2D} \Delta t & -\frac{R}{2D} \Delta t \end{bmatrix}. \tag{10}$$

The global framework of robot is described as shown in Fig. 5.

Then, we can easily obtain the measurement model that shown as

$$h(x) = \begin{bmatrix} \rho \\ \alpha \end{bmatrix} = \begin{bmatrix} \sqrt{(x_l - x)^2 + (y_l - y)^2} \\ \arctan(\frac{y_l - y}{x_l - x}) - (\theta - \frac{\pi}{2}) \end{bmatrix}, \tag{11}$$

Fig. 4 The motion model

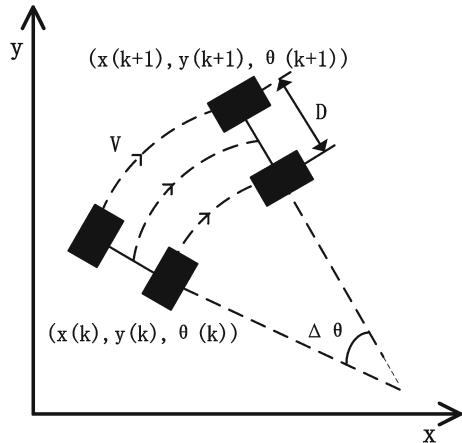
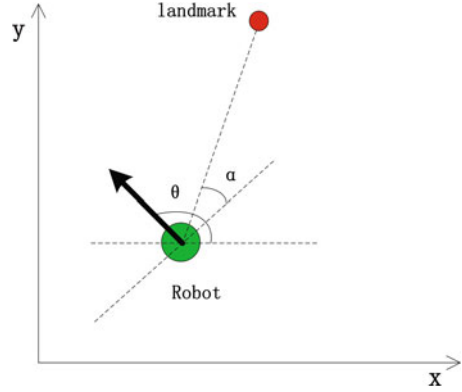


Fig. 5 The global framework of robot



where x_l, y_l are the Cartesian coordinates of landmarks, x and y are the positions of robot, and θ is the heading angle. Similarly, by linearizing the measurement model, we obtain the Jacobian matrix H_x of measurement model that shown as

$$H_x = \begin{bmatrix} \frac{x-x_l}{\rho} & \frac{y-y_l}{\rho} & 0 \\ \frac{y_l-y}{\rho^2} & \frac{x-x_l}{\rho^2} & -1 \end{bmatrix}. \quad (12)$$

3.4 Motion Plan and Obstacle Avoidance

Collision avoidance has been widely studied in autonomous mobile system, and several techniques has been proposed to tackle with collision avoidance. In this paper, we utilize a extended beam curvature method. The details are illustrated as follows:

Obtain the environmental information The laser's range is from 0° to 270° with 1° resolution. As a result, in every step, we obtain 271 values and every value d_i represents the distance of the i th beam.

Evaluate each beam To find the best beam, we evaluate the beams through a equation [11]:

$$f_e(d_i, p_i) = \alpha d_i \cos(|dp_i|) - (1 - \alpha) \left| \frac{dp_i}{\pi} \right|, \quad (13)$$

where p_i is defined as the angle between the X axis and the i th beam and p_G is defined as the angle between the X axis and the goal direction in the coordinates system which is illustrated in Fig. 6. $dp_i = p_G - p_i$, $d_i \cos(|dp_i|)$ may be regarded as the projected distance on the goal direction, and α and β are constants.

Searching the instant goal candidate and safe path In order to find the proper instant goal, it is essential to find the potential safe area. The safe area could be searched by [11]:

$$p_{1\text{safe}} = \max_{(P_i, d_i)} \left(P_i + \frac{d_{\text{safe}}}{d_i} \right), -\pi < P_i < P_{\text{best}}, \tag{14}$$

$$p_{2\text{safe}} = \min_{(P_i, d_i)} \left(P_i - \frac{d_{\text{safe}}}{d_i} \right), P_{\text{best}} < P_i < \pi, \tag{15}$$

where $p_{2\text{safe}}, p_{1\text{safe}}$ are the top angle and the bottom angle of the safe area and d_{safe} is a safe threshold adjusted in experiments. Then, we decide the heading angle through comparing the angle between robot's position and goal's position with the limits of safe area [11]:

$$\theta_{\text{heading}} = \begin{cases} p_0, & p_0 \in [p_{1\text{safe}}, p_{2\text{safe}}] \\ p_{1\text{safe}}, & p_0 < p_{1\text{safe}} \\ p_{2\text{safe}}, & p_0 > p_{2\text{safe}} \end{cases} \tag{16}$$

Determining velocity According to the instant goal and safe path, we determine angular velocity ω such that the heading angle of the robot is approaching the instant goal.

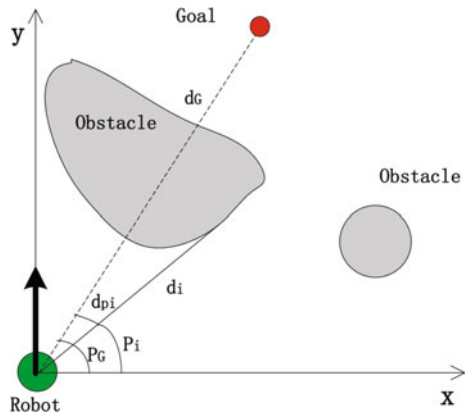
$$\omega = k_\omega (\theta_i - \theta_{\text{heading}}) \tag{17}$$

where k_ω is an adjustable constant, θ_i is the angle between the robot and the instant goal, and θ_{heading} is the heading angle of the robot. Moreover, in order to make sure that the robot will not deviate from the safe path suddenly, it is essential to limit the maximum linear speed v_{MAX} . The linear velocity v can be defined as follows:

$$v = \begin{cases} v_{\text{MAX}} - k_v |\omega| & v > 0 \\ 0 & v < 0 \end{cases} \tag{18}$$

$$v_r = \frac{2v + \omega D}{2} \tag{19}$$

Fig. 6 The beam appraisal



$$v_l = \frac{2v - \omega D}{2}, \quad (20)$$

where v is the linear velocity of the robot and v_r and v_l are the forward velocities of the right and left wheels.

Detecting goal We detect whether robot has arrived the goal position through the distance constraints

$$|x - x_{\text{goal}}| < \sigma_e, \quad (21)$$

$$|y - y_{\text{goal}}| < \sigma_e, \quad (22)$$

where $x_{\text{goal}}, y_{\text{goal}}$ are the position of goal and σ_e is the error tolerance.

4 Experiments

In the experiments, the parameters of the robot are set as $R = 0.10$ m, $D = 0.6$ m, $k_v = 1.0$, $k_\omega = 0.4$, $\alpha = 0.05$, $\beta = 0.95$, and $\sigma_e = 0.2$ m, and the sampling interval is $\Delta T = 0.1$ s.

4.1 Estimate Start Position

Placing the robot at different initial positions, we begin to estimate the start position, and the result is shown in Table 1.

In the table, x, y are the practical positions, x_e, y_e are the estimated positions, and e_x, e_y are the relative errors. From the table, we can see that the results are acceptable for extended Kalman filter algorithm.

4.2 Obstacle Avoidance

The pulse number per circle is 2048, and if the control input is angular velocity, we should turn it into angular velocity. The laser's accuracy is 0.01 m and 1° .

Table 1 Estimated result

x	x_e	$e_x(\%)$	y	y_e	$e_y(\%)$
-5.4	-5.4865	1.601	0.92	0.9665	5.054
-5.4	-5.4738	1.367	1.20	1.2882	6.846
-4.8	-4.8729	1.518	1.20	1.2715	5.958
-4.8	-4.8830	1.729	0.60	0.6429	7.150

Fig. 7 The built map of SLAM

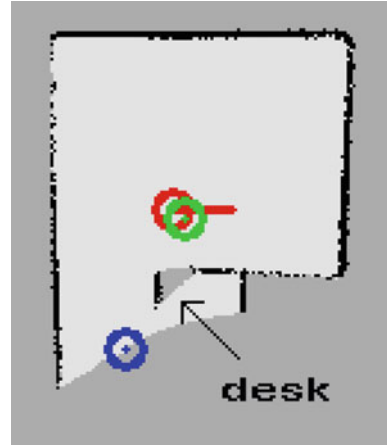
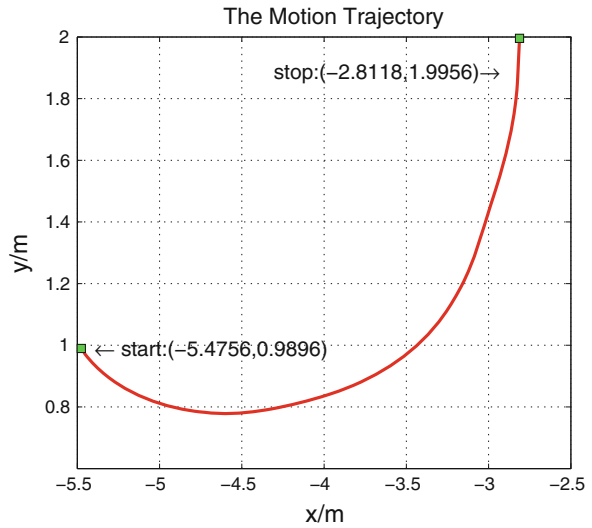


Fig. 8 The motion trajectory of robot



So we initialize the measurement noise covariance matrix Q and control noise covariance matrix R as follows:

$$R = \begin{bmatrix} (1.0/2048/\pi/\Delta T)^2 & 0 \\ 0 & (1.0/2048/\pi/\Delta T)^2 \end{bmatrix}, \quad Q = \begin{bmatrix} (0.01)^2 & 0 \\ 0 & (1.0 * \pi/180)^2 \end{bmatrix}.$$

After initialized these matrices, we place the robot at $(-5.4, 1.0)$ and set the goal position at $(-3.0, 2.1)$ and then run our program. Finally, we obtain the map of our experimental environment as shown in Fig. 7.

In Fig. 7, the blue and green circles are the original position and the goal position, respectively. The red circle represents the robot position. The motion

trajectory is shown in Fig. 8. During the overall process, robot moves toward left to acquire the enough diameter to bypass the desk on the right. Then, robot turns right to avoid the desk and moves to the goal. From the motion trajectory and the velocity trajectory, we can see that the whole process is smooth and continuous and the final errors are 0.1882 and 0.1044 m, which are within σ_e .

5 Conclusions

In this paper, a mobile system which integrates SLAM and obstacle avoidance is implemented. In SLAM, through comparing the detected features with landmarks which set previously in global coordinate system, we utilize EKF to estimate the pose of robot. Meanwhile, an obstacle avoidance algorithm is planning the motion path to ensure that the robot will safely arrive the goal position. In the implementation, the results are satisfying.

Acknowledgments This work is supported by the Natural Science Foundation of China under Grants 61174045 and 61111130208, and the International Science and Technology Cooperation Program of China under 2011DFA10950, and the Fundamental Research Funds for the Central Universities (No. 2011ZZ0104) and the Program for New Century Excellent Talents in University No. NCET-12-0195.

References

1. Kalman R (1960) A new approach to linear filtering and prediction problems. *Trans ASME J Basic Eng* 82:35–45
2. Dellaert F, Fox D, Burgard W, Thrun S (1999) Monte Carlo localization for mobile robots. *International conference on robotics and automation*, vol 2, pp 1322–1328
3. Hu HS (2000) Landmark based navigation of autonomous robots in industry. *Int J of Ind Robot* 27:458–467
4. Roumeliotis S, Bekey G (2000) Bayesian estimation and Kalman filtering: a unified framework for mobile robot localization. In: *Proceedings of 2000 IEEE international conference on robotics and automation*, vol 23, pp 2985–2992
5. Smith RC, Self M, Cheeseman P (1986) Estimating uncertain spatial relationships in robotics. In: *Proceedings of the 2nd annual conference on uncertainty in, Artificial Intelligence (UAI-86)*, pp 267–288
6. Keron Y, Borenstein J (1991) Potential field methods and their inherent limitations for mobile robot navigation. In: *Proceedings of the IEEE conference on robotics and automation*, vol 2, pp 1398–1404
7. Glasius R, Komoda A, Gielen S (1995) Neural network dynamics for path planning and obstacle avoidance. *Neural Networks* 8:125–133
8. Ulrich I, Borenstein J (2000) VFH*: local obstacle avoidance with look-ahead verification. In: *IEEE international conference on robotics and automation*, vol 3, pp 2505–2511
9. Simmons R (1996) The curvature-velocity method for local obstacle avoidance. In: *Proceedings of the 1996 IEEE international conference on robotics and automation Minneapolis*, vol 4, pp 3375–3382

10. Ko N, Simmons G (1998) The lane-curvature method for local obstacle avoidance. In: Proceedings of the IEEE/RSJ international conference on intelligent robots and systems, vol 3, pp 1615–1621
11. Fernández JL, Sanz R, Benayas JA, Díguez AR (2005) Improving collision avoidance for mobile robots in partially known environments: the beam curvature method. *Rob Auton Syst* 46:205–219
12. Nguyen V, Gächter S, Martinelli A, Tomatis N, Siegwart R (2007) A comparison of line extraction algorithms using 2D range data for indoor mobile robotics. *Auton Robot* 23:97–111
13. Pavlidis T, Horowitz SL (1974) Segmentation of plane curves. *IEEE Trans Comput C-* 23:860–870
14. Kuo BW, Chang HH, Chen YC, Huang SY (2011) A light-and-fast SLAM algorithm for robots in indoor environments using line segmentMap. *J Robot*
15. Li Z, Xie Z, Ming A (2008) Simultaneously firing sonar ring based high-speed navigation for non-holonomic mobile robots in unstructured environment. *Int J Veh Auton Syst* 6:172–185
16. Zhu X, Li Z, Ye W, Zhao H (2001) Event-based Tele-presence of a mobile service robot. In: IEEE symposium on haptic audio visual environments and games, pp 118–123

A Novel Approach for Computing Quality Map of Visual Information Fidelity Index

Yu Shao, Fuchun Sun, Hongbo Li and Ying Liu

Abstract The visual information fidelity (VIF) index gained widespread popularity as a tool to assess the quality of images and to evaluate the performance of image processing algorithms and systems. But VIF is not a map-based quality metric if its quality map is calculated by traditional sliding window approach. This map-based property is owned by the other quality metrics such as structural similarity (SSIM) and mean-squared error (MSE). In this article, we first construct a novel VIF quality map in pixel domain, which makes VIF become a Minkowski norm of its quality map. Furthermore, we deduce the gradient of VIF by taking the derivative of VIF index with respect to the reference image. The gradient of VIF is easy to calculate and has many useful applications. Experimental results show that the proposed quality map can provide useful guidance on how local image quality is similar to reference image.

Keywords Visual information fidelity · Quality map · Structural similarity · Image quality assessment

1 Introduction

Image quality assessment (IQA) is a fundamental issue in many image processing applications. It can be applied to optimize and design the image processing systems and algorithms to improve the visual quality. For an IQA method, its quality map and gradient are two most important attributes. Quality map measures image quality locally and is able to capture local dissimilarities when compared to reference image. In quality map image, the brighter regions mean better quality.

Y. Shao (✉) · F. Sun · H. Li · Y. Liu

State Key Laboratory of Intelligence Technology and Systems, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, People's Republic of China
e-mail: shaoyu2011@foxmail.com

Traditionally, MSE and SSIM have been widely used not only for the evaluation of image quality, but also for the design and optimization of signal processing algorithms and systems [1]. That is because their quality map and gradient can be easily computed. For instance, the SSIM index [2, 3] is computed locally at each pixel of the image and can be visualized as an image, often referred to as a SSIM map, which provides useful information on the localization of distortions. The gradient of SSIM with respect to the image has been derived in [4–6], and this gradient is used as a fidelity term in iterative optimization procedures.

The VIF is the most accurate image quality metric according to the performance evaluation of major image quality assessment algorithms performed in [7]. In spite of its high level of accuracy, this index has not been given as much consideration as the SSIM index in a variety of applications, far behind its widespread usage as purely an assessment or comparison tool in other applications. There has been also a growing interest of using VIF as an objective function in optimization problems in a variety of image processing applications [8]. One major problem that could strongly impede the progress of further applications is the lack of understanding and desirable mathematical properties of VIF. For example, the VIF is a nonmap-based quality metric which gives a final score for a distortion image. Unlike the SSIM and MSE which compute a quality (or distortion) map between the reference and distorted images to depict the distribution of quality degradation at image pixels, the overall quality is usually computed as a mean over all the pixels in the distortion map. Second, SSIM or MSE methods can easily calculate its gradient. However, due to the high computational complexity of VIF (6.5 times the computation time of the SSIM index according to [9]) and its nonmap-based quality metric character, it is hard to get the gradient of VIF.

Some researchers are trying to study these two essential properties of VIF. For instance, Seshadrinathan [10] analyzed the properties of SSIM and VIF and established a relationship between SSIM and VIF. Li [11] proposed a spatial information theoretical weighting map for SSIM and VIF. Brighter regions in this weighting map indicate larger weights during error pooling process of IQA. But these studies are only qualitative research. According to the best of our knowledge, there is no complete formula about the quality map and gradient of VIF. In this article, we first propose a novel quality map for VIF in pixel domain. We convert the VIF from a nonmap-based quality metric to a map-based quality metric. Experimental results show that the proposed VIF quality map provides useful guidance on how local image quality is similar to reference image. Based on the formulation of VIF quality map, we then deduce the gradient of VIF by taking the derivative of VIF with respect to reference image. This gradient can be used to solve the minimization problems with VIF term in image processing field.

This paper is organized as follows. The principle of VIF is introduced in Sect. 1. Section 3 presents the proposed quality map of VIF, and its gradients are discussed in Sect. 4. The experiments are analyzed in Sect. 5, and conclusions are drawn in Sect. 6.

2 The VIF Metric

The VIF proposed by Sheikh [9] is an IQA method that consistently outperforms almost all other approaches. It treats the IQA as an information fidelity problem based on natural scene statistics (NSS) theory. There are two types of VIF: wavelet domain version and pixel domain version. Considering that wavelet domain version VIF is more complex and our proposed quality map is based on pixel domain, we only discuss the pixel domain version of VIF.

Support C and D denote the random fields (RFs) from the reference and distorted images, respectively. Let $C^N = (C_1, C_2, \dots, C_N)$ and denote N elements from C , and let $D^N = (D_1, D_2, \dots, D_N)$ be the corresponding N elements from D . C is a product of two stationary RFs that are independent of each other:

$$C = S \cdot U = \{S_k \cdot U_k : k \in I\}, \quad (1)$$

where I denotes the set of spatial indices for the RFs, S is an RFs of positive scalars, and U is a Gaussian scalar RFs with mean zero and variance σ_U^2 . The image distortion model is a signal attenuation and additive Gaussian noise, defined as follows:

$$D = GC + V = \{g_k C_k + V_k : k \in I\}, \quad (2)$$

where G is a deterministic scalar attenuation field and V is a stationary additive zero-mean Gaussian noise RFs with variance σ_V^2 .

The human visual system (HVS) model in VIF quantifies the impact of the image that flows through HVS:

$$\begin{aligned} E &= C + N, \\ F &= D + N, \end{aligned} \quad (3)$$

where E and F denote the cognitive output of the reference and test images extracted from the brain, respectively; N represents stationary, white Gaussian noise RFs with variance σ_n^2 .

VIF utilizes mutual information $I(C_k, E_k)$ to measure the information that can be extracted from the output of HVS when the reference image is being viewed:

$$I(C_k, E_k) = \frac{1}{2} \log_2 \left(\frac{|s_k^2 C_U + \sigma_N^2 I|}{|\sigma_N^2 I|} \right) = \frac{1}{2} \log_2 \left(1 + \frac{\sigma_C^2}{\sigma_N^2} \right). \quad (4)$$

In addition, information $I(C_k, F_k)$ is measured in the same way when the test image is being viewed:

$$I(C_k, F_k) = \frac{1}{2} \log_2 \left(\frac{|g_k^2 s_k^2 C_U + (\sigma_N^2 + \sigma_{V_k}^2) I|}{|(\sigma_N^2 + \sigma_{V_k}^2) I|} \right) = \frac{1}{2} \log_2 \left(1 + \frac{g_k^2 \sigma_C^2}{\sigma_N^2 + \sigma_{V_k}^2} \right). \quad (5)$$

The above mutual information assumes that the distortion model parameters g and σ_V^2 are known a priori, but these would need to be estimated in practice. An estimated model replaces the theoretical model in practice. The value of the filed g over block k is denoted as g_k , and the variance of the RFs V over block k is denoted as σ_{V_k} , both are estimated from the local variance of pixels based on maximum likelihood (ML) criteria, which are easily estimated by

$$\hat{g} = \sigma_{CD} / \sigma_C^2, \quad (6)$$

$$\hat{\sigma}_V^2 = \sigma_D^2 - \hat{g}\sigma_{CD}, \quad (7)$$

where

$$\begin{aligned} \mu_C &= w * C, \\ \mu_D &= w * D, \\ \sigma_C^2 &= w * (C - \mu_C)^2 = w * C^2 - \mu_C^2, \\ \sigma_D^2 &= w * (D - \mu_D)^2 = w * D^2 - \mu_D^2, \\ \sigma_{CD} &= w * (C - \mu_C)(D - \mu_D) = w * (CD) - \mu_C\mu_D, \end{aligned} \quad (8)$$

in which w is a symmetric low-pass kernel (e.g., 11×11 normalized Gaussian kernel). ‘*’ denotes convolution.

Sheikh [9] uses a more sophisticated vector GSM model for VIF. Rezazadeh [12] uses scalar GSM instead of vector GSM in modeling the images for VIF computation. Rezazadeh [13] shows that the first-level approximation subband of decomposed images plays an important role in improving quality assessment performance and also in complexity reduction. So we restrict our analysis to a scalar version of the VIF metric, where the natural scene model is identical to that used in the scalar IFC (information fidelity criterion) [14] index.

Considering a single subband, we obtain the sample VIF as follows:

$$\text{VIF}(C, D) = \frac{I(C^N, F^N)}{I(C^N, E^N)} = \frac{\sum_k I(C_k, F_k)}{\sum_k I(C_k, E_k)} = \frac{\sum_{k=1}^N \log_2 \left(1 + \frac{g_k^2 \sigma_{C_k}^2}{\sigma_N^2 + \sigma_V^2} \right)}{\sum_{k=1}^N \log_2 \left(1 + \frac{\sigma_{C_k}^2}{\sigma_N^2} \right)}. \quad (9)$$

The denominator of the above equation represents the amount of information that the HVS can extract from the original image. The numerator represents the amount of information that the HVS can extract from the distorted image. The ratio of these two quantities hence is a measure of the amount of information in the distorted image relative to the reference image and has been shown to correlate very well with visual quality. The one extra parameter in this model namely the variance of the neural noise σ_N^2 is hand-optimized in [9] and chosen to be 2.

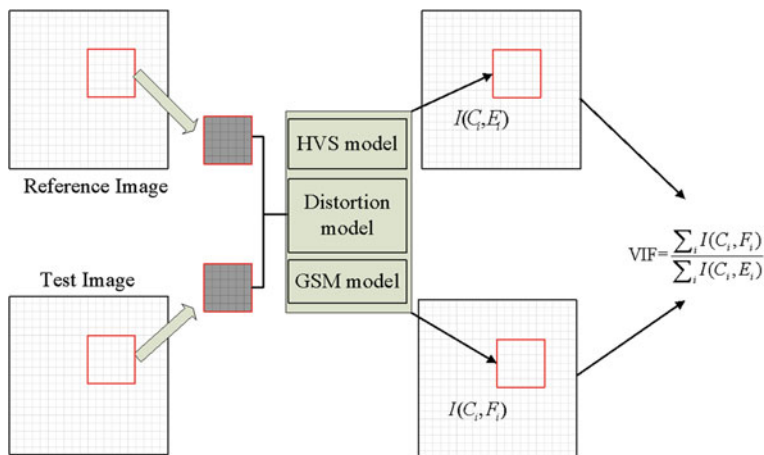


Fig. 1 A schematic of VIF

As depicted in Fig. 1, VIF first decomposes the image into several blocks. Then, VIF measures the visual information by computing mutual information in the different models in each block by Eqs. (4) and (5). Finally, the image quality value is measured by integrating visual information for all the blocks by Eq. (9).

3 Quality Map of VIF

There is no doubt that quality maps are important to the IQA method. Different image quality maps can provide a substantially distinct prediction of local image quality. Although a lot of research effort has been put into investigating the perceptual error map, such as the absolute difference map and the SSIM map, much less has been done for studying the perceptual error map or quality map of VIF. In the former case[15], the VIF is only one number that quantifies the information fidelity for the entire image, whereas in the latter case, a sliding window approach could be used to compute a quality map that could visually illustrate how the visual quality of the test image varies over space. However, in contrast to most previous quality assessment methodologies, the VIF is not a Minkowski norm of the quality map. In other words, a norm of the VIF quality map is not the appropriate measure of image quality.

On the right side of the Eq. (9), the numerator is basically IFC and the denominator can be thought as a content-dependent adjustment. For reference image \mathbf{X} and test image \mathbf{Y} , we define the numerator $R(\mathbf{X}, \mathbf{Y}, i, j)$ and denominator $P(\mathbf{X})$ as follows:

$$R(\mathbf{X}, \mathbf{Y}, i, j) = \log_2 \left(1 + \frac{g_k^2 \sigma_{X_k}^2}{\sigma_N^2 + \sigma_V^2} \right), \quad (10)$$

$$P(\mathbf{X}) = \sum_{k=1}^N \log_2 \left(1 + \frac{\sigma_{X_k}^2}{\sigma_N^2} \right) \quad (11)$$

here, N is the number of pixels in either of the input images.

At each point k with its coordinate (i, j) , VIF_{map} is an indication of the local similarity between reference image \mathbf{X} and test image \mathbf{Y}

$$\text{VIF}_{\text{map}}(\mathbf{X}, \mathbf{Y}, i, j) = \frac{R(\mathbf{X}, \mathbf{Y}, i, j)}{P(\mathbf{X})}. \quad (12)$$

The quality map $\text{VIF}_{\text{map}}(\mathbf{X}, \mathbf{Y}; i, j)$ is then added up to obtain a single quality score for the entire image. The VIF for an image \mathbf{Y} , with respect to the reference image \mathbf{X} , is given by the following equation:

$$\text{VIF}(\mathbf{X}, \mathbf{Y}) = \sum_{\forall i, j} \text{VIF}_{\text{map}}(\mathbf{X}, \mathbf{Y}, i, j). \quad (13)$$

This VIF is a Minkowski norm of its quality map. So we convert the VIF from a nonmap-based quality metric to a map-based quality metric. Though our proposed approach for computing VIF quality map has a form that is more complicated than that of SSIM, it still remains analytically tractable as discussed in subsequent sections. Specifically, we can deduce the deviation of VIF. This makes VIF being able to produce a spatially varying quality map in which quality varies across the image. So the final output of the VIF is either a spatial map showing the image quality at different spatial locations or a single number describing the overall quality of the image.

4 Gradient of VIF

Based on our formulation of VIF and its quality map given above, we first compute $\partial \text{VIF} / \partial Y(a, b)$ and then $\nabla_{\mathbf{Y}} \text{VIF}(\mathbf{X}, \mathbf{Y})$.

$$\frac{\partial}{\partial Y(a, b)} \text{VIF} = \sum_{\forall i, j} \frac{\partial}{\partial Y(a, b)} \text{VIF}_{\text{map}} \quad (14)$$

in which

$$\frac{\partial}{\partial Y(a, b)} \text{VIF}_{\text{map}} = \frac{\partial \sigma_{XY}}{\partial Y(a, b)} \frac{\partial \text{VIF}_{\text{map}}}{\partial \sigma_{XY}} + \frac{\partial \sigma_Y^2}{\partial Y(a, b)} \frac{\partial \text{VIF}_{\text{map}}}{\partial \sigma_Y^2}. \quad (15)$$

By calculating partial derivatives of the parameters defined in Eq. (8) with respect to $Y(a, b)$, we have

$$\frac{\partial \sigma_{XY}}{\partial Y(a, b)} = \omega(i - a, j - b)(X(a, b) - \mu_X), \quad (16)$$

$$\frac{\partial \sigma_Y^2}{\partial Y(a, b)} = 2\omega(i - a, j - b)(Y(a, b) - \mu_Y). \quad (17)$$

By substituting the partial derivatives in Eq. (15) and collecting $\omega(i - a, j - b)$, the summation in Eq. (14) turns into a weighted sum of three convolutions:

$$\nabla_Y \text{VIF}(\mathbf{X}, \mathbf{Y}) = w * \mathbf{M}_1 + \left(w * \frac{\partial \text{VIF}_{\text{map}}}{\partial \sigma_{XY}} \right) \mathbf{X} + \left(w * \frac{\partial \text{VIF}_{\text{map}}}{\partial \sigma_Y^2} \right) \mathbf{Y} \quad (18)$$

where the simplified auxiliary variable \mathbf{M}_1 is

$$\mathbf{M}_1 = -\mu_X \frac{\partial \text{VIF}_{\text{map}}}{\partial \sigma_{XY}} - 2\mu_Y \frac{\partial \text{VIF}_{\text{map}}}{\partial \sigma_Y^2}. \quad (19)$$

Using Eq. (18), we can compute $\nabla_Y \text{VIF}(\mathbf{X}, \mathbf{Y})$ for all pixels with just three convolutions and some element-wise multiplications and additions. The partial derivatives required in Eq. (19) are given below:

$$\frac{\partial \text{VIF}_{\text{map}}}{\partial \sigma_{XY}} = \frac{1}{P} \left(1 + \frac{g^2 \sigma_X^2}{\sigma_V^2 + \sigma_n^2} \right)^{-1} \frac{2g(\sigma_V^2 + \sigma_n^2)\sigma_X^2 + 2g^2 \sigma_{XY}}{(\sigma_V^2 + \sigma_n^2)^2} \quad (20)$$

$$\frac{\partial \text{VIF}_{\text{map}}}{\partial \sigma_Y^2} = \frac{1}{P} \left(1 + \frac{g^2 \sigma_X^2}{\sigma_V^2 + \sigma_n^2} \right)^{-1} \frac{-g^2 \sigma_X^2}{(\sigma_V^2 + \sigma_n^2)^2} \quad (21)$$

in which P is defined in Eq. (11).

Since VIF can be employed as the data-fidelity term of optimization function in some image processing field, if having the gradient of VIF, the gradient descent approach can be used with an iterative procedure to solve the optimization problem.

5 Experimental Results

Like SSIM map, our proposed VIF map is computed locally at each pixel of the distorted image and can be visualized as an image, which provides useful information on the location of distortions. To validate the proposed VIF map model, we

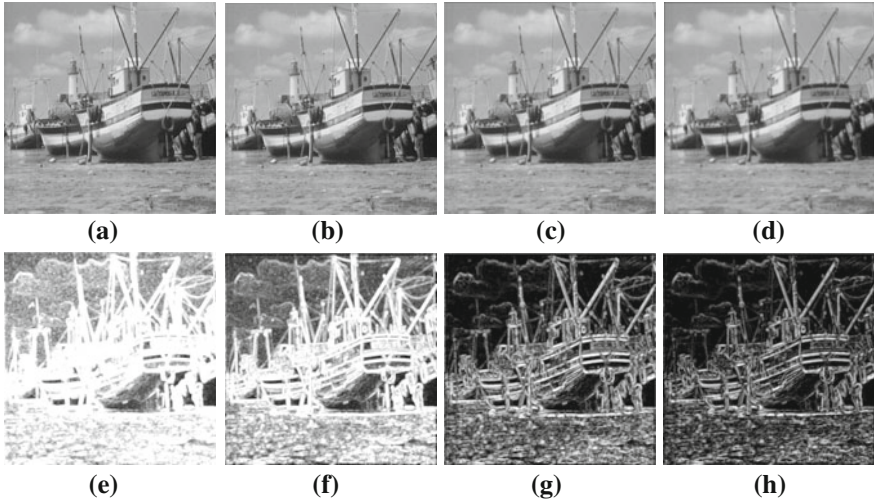


Fig. 2 Illustration the VIF map of distorted images with different blur levels. **a** Original image(VIF = 1). **b** Blurred image,Gaussian blur kernelsize = 7×7 and $\sigma = 0.5$ (VIF = 0.674). **c** Blurred image,Gaussian blur kernelsize = 7×7 and $\sigma = 1$ (VIF = 0.293). **d** Blurred image,Gaussian blur kernelsize = 7×7 and $\sigma = 1.5$ (VIF = 0.195). **e** VIF map of (a). **f** VIF map of (b). **g** VIF map of (c). **h** VIF map of (d)

first test its performance using two types of image distortion: Gaussian blur and additive white Gaussian noise.

Figure 2 illustrates the VIF map of distorted images with different Gaussian blur levels. First row shows the reference image (a) and distorted images (b–d) obtained from the reference using Gaussian blur with an incremental variance. Second row shows the corresponding VIF maps at each pixel displayed as an image. Figure 3 illustrates the VIF map of distorted images with different noise levels. First row shows the reference image (a) and distorted images (b–d) obtained from the reference using additive, white Gaussian noise with an incremental noise level. Second row shows the corresponding VIF maps. In the VIF map image, bright regions correspond to better quality and dark regions correspond to worst quality. From Figs. 2 and 3, we obviously see that with noise or blur level increase, the visual quality of distortion image deteriorates gradually, and its quality map brightness also reduces gradually. The proposed VIF map clearly displays the regions of the distorted image that are visually annoying to the human observer.

Figure 4 shows a reference image that has been distorted with three different types of distortion and its corresponding VIF map. The distortion types illustrated are contrast stretch, Gaussian blur, and JPEG compression. VIF map of Fig. 4e–h shows the spread of structural information. In flat image regions such as the sky area, the information content of the image is low, whereas in textured regions and regions containing strong edges such as the outline of buildings, the image quality is high. The contrast-enhanced image Fig. 4b has a brighter quality map than

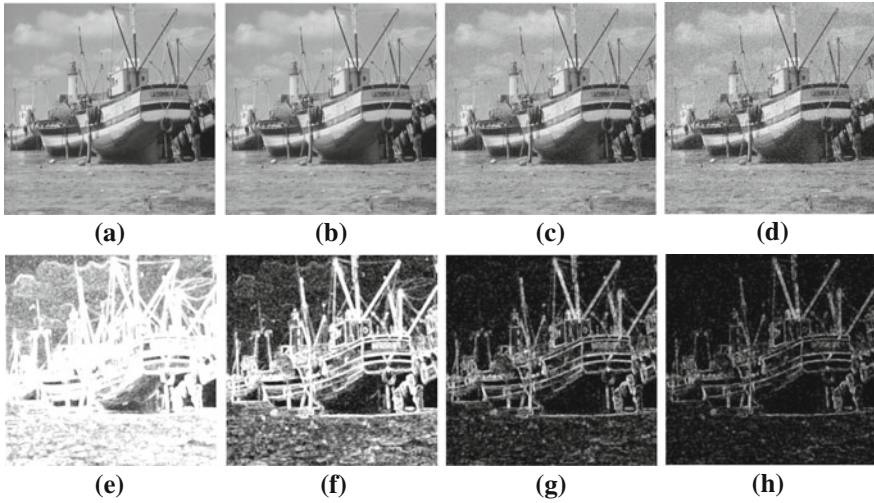


Fig. 3 Illustration the VIF map of distorted images with different noise levels. **a** Original image(VIF = 1). **b** Noised image,Gaussian whitenoise with $\sigma = 5$ (VIF = 0.442). **c** Noised image,Gaussian whitenoise with $\sigma = 15$ (VIF = 0.156). **d** Noised image,Gaussian whitenoise with $\sigma = 25$ (VIF = 0.083). **e** VIF map of (a). **f** VIF map of (b). **g** VIF map of (c). **h** VIF map of (d)

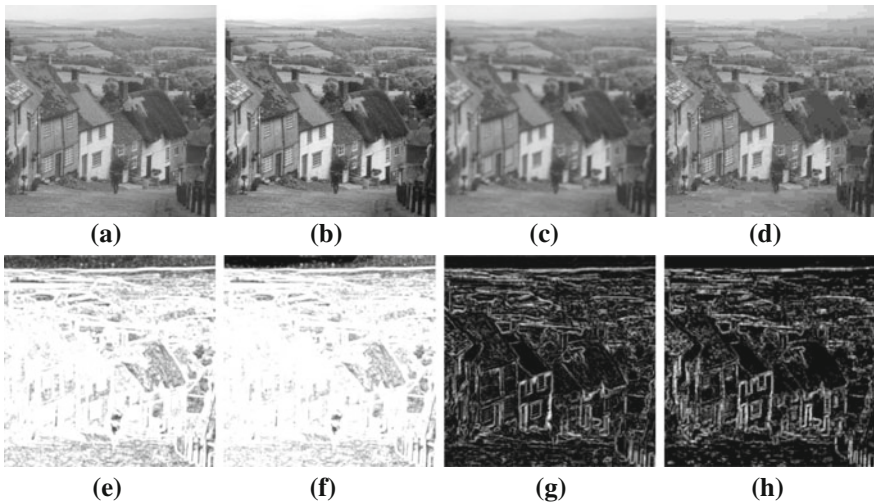


Fig. 4 The VIF map can capture the quality loss or improvement in the distorted image. **a** reference image (VIF = 1), **b** distorted image by contrast stretch (VIF = 1.1), **c** distorted image by Gaussian Blur (VIF = 0.07), **d** distorted image by JPEG compression (VIF = 0.10), **e-h** are corresponding VIF map of image **a-d**.

reference image Fig. 4a, and its VIF value is larger than unity. In contrast, both the blurred image Fig. 4c and the JPEG-compressed image Fig. 4d have a darker quality map compared with the reference image, and its VIF value is smaller than unity. It is interesting to see that the brightness of VIF map reflects its VIF value. It indicates the relative image information that is present in the distorted image. So this proposed VIF map captures the improvement or loss of the distorted image in visual quality and appears to be a good indicator of image quality in the pixel domain.

6 Conclusion

The goal of this paper was to analyze the properties of recently most widely used IQA paradigm VIF based on information theoretical framework. We first showed that the numerator term in VIF, which is the mutual information between the test image and the reference image, can be interpreted as a quality map for VIF. The experimental results indicate that the proposed VIF map provides useful guidance on how local image quality is similar to reference image. Additionally, VIF map can also predict improvement in quality over space. Based on the formulation of VIF map, we then deduce the gradient of VIF by taking the derivative of VIF with respect to reference image. We pointed out that this gradient can be used to solve optimization problem where there exists VIF term. This VIF value is the sum of VIF map. In the future, we would like to use other pool strategies to compute the overall VIF value and extend our analysis to the multi-subband/vector VIF model.

Acknowledgments This study was jointly funded by the National Basic Research Program of China (Grant No:2012CB821206) and the Tsinghua Self-innovation Project (Grant No:20111081111).

References

1. Wang Z, Bovik AC (2009) Mean squared error: love it or leave it? a new look at signal fidelity measures. *IEEE Sign Process Mag* 26:98–117
2. Wang Z, Simoncelli EP (2004) Stimulus synthesis for efficient evaluation and refinement of perceptual image quality metrics. *Human vision and electronic imaging IX. Proc SPIE* 5292:99–108
3. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Proc* 13:600–612
4. Avanaki AN (2009) Exact global histogram specification optimize for structural similarity. *Opt Rev* 16:613–C621
5. Rehman A, Rostami M, Wang Z, Brunet D, Vrscay ER (2012) SSIM-inspired image restoration using sparse representation. *EURASIP J Adv Signal Proc* 2012:16
6. Wang Z, Simoncelli EP (2008) Maximum differentiation (MAD) competition a methodology for comparing computational models of perceptual quantities. *J Vision* 8 8:1–13

7. Sheikh HR, Sabir MF, Bovik AC (2006) A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Trans Image Proc* 15:3440–3451
8. Seshadrinathan K, Bovik AC (2007) New vistas in image and video quality assessment. *Proc SPIE* 6492:649202
9. Sheikh HR, Bovik AC (2006) Image information and visual quality. *IEEE Trans Image Proc* 15:430–444
10. Seshadrinathan K, Bovik AC (2008) Unifying analysis of full reference image quality assessment. In: 15th IEEE international conference on image processing, San Diego, CA, United states pp 1200–1203
11. Li Q (2009) Objective image and video quality assessment with applications. PhD thesis, The University of Texas at Arlington
12. Rezazadeh S, Coulombe S (2011) A novel discrete wavelet transform framework for full reference image quality assessment. *Sign Image Video Proc* pp 1–15
13. Rezazadeh S, Coulombe S (2010) Low-complexity computation of visual information fidelity in the discrete wavelet domain. In: IEEE international conference on acoustics speech and signal processing (ICASSP) 2010:2438–2441
14. Sheikh HR, Bovik AC, de Veciana G (2005) An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Trans Image Proc* 14:2117–2128
15. Sheikh HR (2004) Image quality assessment using natural scene statistics. PhD thesis, The University of Texas at Austin

A Simulation of Electromagnetic Compatibility QoS Model in Cognitive Radio Network

Hai-yu Ren, Ming-xue Liao, Xiao-xin He and Kun Xu

Abstract Cognitive radio system is a typical cognitive system. Spectrum decision in cognitive radio must consider the QoS requirement. The requirements in this paper are to ensure the largest network and the optimal parameter for communication at the same time. To satisfy these requirements, spectrum decision needs to select appropriate communication parameters, adjust subnet topology, and take electromagnetic compatibility (EMC) into account for multi-transceiver users. This paper analyzes the principle of electromagnetic compatibility and gives the flow of electromagnetic compatibility analysis (EMCA). Then, a precise QoS model based on the EMCA is proposed. The precise EMC QoS (pEMC) model takes the topological bottlenecks, electromagnetic compatibility, and spectrum pool capacity into account. The complexity of EMCA will rapidly increase with the spread of spectrum pool of topology, and a real-time spectrum decision cannot be guaranteed. Then, an approximate EMC QoS (aEMC) model is proposed and the consistency of these two models is simulated. The results show a high consistency between the pEMC and aEMC model so that the aEMC model can substitute for the pEMC model in practice.

H. Ren (✉) · M. Liao · X. He · K. Xu

State Key Laboratory of Integrated Information System Technology, Institute of Software,
Chinese Academy of Sciences, Beijing, China
e-mail: haiyuren@gmail.com

M. Liao

e-mail: mingxue@iscas.ac.cn

X. He

e-mail: xiaoxin@iscas.ac.cn

K. Xu

e-mail: xukun10@iscas.ac.cn

K. Xu

University of Chinese Academy of Sciences, Beijing, China

Keywords Cognitive radio · Spectrum decision · Electromagnetic compatibility · QoS

1 Introduction

Joseph. Mitola first proposed the concept of cognitive radio (CR) in 1999 [1, 2]. Soon, FCC and Simon Haykin defined the CR, respectively [3, 4]. In 2005, Simon Haykin proposed a basic cognitive cycle. He considered the CR as a feedback communication system and the cycle consisted of three processes including spectrum sensing, spectrum analysis, and spectrum decision.

Spectrum decision has a tight relationship with QoS in different application scenes [5–7]. The QoS requirements in this paper are to ensure a largest network and a best parameter for communication at the same time. A largest network means that the topology includes more child secondary users (*CSU*) and a best parameter means that the topology has most many frequency combinations. These two QoS requirements interrelate to each other but restrict themselves to each other also.

At the same time, a secondary cognitive user who has several transceivers needs to make electromagnetic compatible (EMC). An analysis on EMC among several transceivers is called a process of EMCA, which determines whether interferences such as harmonic interference, intermodulation interference caused by non-linear mixing or co-channel/adjacent channel interference which cannot be filtered, exist.

In Sect. 2, our application scene is introduced. In Sect. 3, the EMCA is described in detail. In Sect. 4, the precise EMC QoS (pEMC) model is proposed to justify the optimal topology. As the complexity of EMC analysis in pEMC model is tightly related with the number of usable frequencies between the parent secondary user (*PSU*) and a subnet of its *CSUs*, a real-time spectrum decision based on pEMC model cannot be guaranteed. Then, an approximate EMC (aEMC) model is proposed. In Sect. 5, three simulations are designed. The EMC probability is simulated in different-sized subnets and different topologies. The consistency between the pEMC model and the aEMC model is simulated and the feasibility of aEMC model is validated in Sect. 6.

2 The Application Scene and Topology Structure

There are many different topologies in a subnet. The i th topology T_i can be expressed by a set of clusters as (1). A cluster C_j can be expressed by a set of *CSUs* as (2), and the number of *CSUs* decides the cluster's size.

$$T_i = \{C_j | j \in (1, \dots, J)\} \tag{1}$$

$$C_j = \{CSU_k | k \in (1, \dots, K)\} \tag{2}$$

In Fig. 1, a subnet is composed of *PSU*, *CSU*₁, and *CSU*₂, and topology $T_1 = \{C_1, C_2\}$ with cluster $C_1 = \{CSU_1\}$, $C_2 = \{CSU_1, CSU_2\}$.

In a *n-CSU* subnet, the *PSU* obtains Fb_k ($k = 1, 2, \dots, N$) by a spectrum-sensing process, the frequency set usable for *PSU* with *CSU*_{*k*}. The spectrum pool of cluster C_j is F_j as (3) and the spectrum pool of topology T_i is expressed as (4):

$$F_j = \bigcap_{k \in (1, \dots, N)} Fb_k \tag{3}$$

$$\{F_j | j \in (1, \dots, N)\} \text{ short as } \{F_j\}. \tag{4}$$

3 The Principle of Electromagnetic Compatibility Analysis

When several transceivers work together or near to each other, the electromagnetic compatibility analysis (EMCA) must be made. In this section, an antenna-feeder system between two transceivers is proposed first. Then, the co-channel/adjacent channel interference caused by transmitter and harmonic interference, intermodulation interference caused by the non-linear mixing in transmitter or receiver are considered, respectively.

In this section, we assume that transceiver *A*, *B*, *C*, and *D* in the same car work on *fa*, *fb*, *fc*, and *fd*, respectively, and each transmitting power is *P*. The antenna-feeder systems' attenuations are L_{AB} , L_{AD} , L_{BC} , and L_{BD} , and so on.

3.1 Attenuation of Antenna-Feeder System

The attenuation of an antenna-feeder system is denoted as *L* in (5). In Fig. 2, a sketch map is shown.

$$L = L_v + L_{ft} + L_{fr} \tag{5}$$

Fig. 1 Topology and its cluster in a subnet

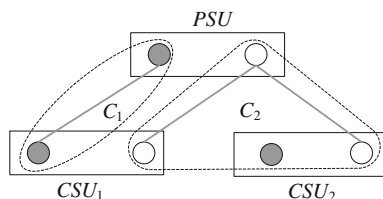
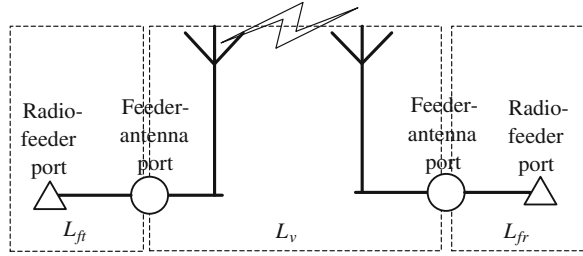


Fig. 2 Antenna-feeder system



L_v is the isolation between transmitter and receiver antennas. L_{ft} and L_{fr} are the feeder attenuations of transmitter and receiver, respectively. L_v (dB) is expressed as

$$L_v = 22 + 20 \times \lg(d/\lambda) - (G_1 + G_2) - (S_1 + S_2). \quad (6)$$

λ (m) is the wavelength of carrier, d (m) is the horizontal distance between antennas, G_1 (dBi) and G_2 (dBi) are the directions of maximum radiation gain, and S_1 (dBp) and S_2 (dBp) are the 90° to the directions of sidelobe levels. In this paper, we suppose that omnidirectional antennas are used, so $S_1 = 0$ and $S_2 = 0$.

3.2 Co-channel/Adjacent Channel Interference

Unwanted signal's frequency in or near the RF bandwidth may cause interference after it is down-converted to the IF. Co-channel interference cannot be avoided. Adjacent channel interference is related to the transmit filter and the IF filter in receiver [8].

3.3 Harmonic Interference

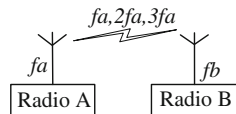
Signal output from non-linear devices contains not only the input signal but also twice or higher-order signals. Such distortion is called as harmonic interference. Generally, only second-order and third-order harmonic interferences must be considered in EMCA [9] (Fig. 3).

Assume that transceiver A can suppress second-order and third-order harmonic R_{sth2} and R_{sth3} , respectively. L is the attenuation of antenna-feeder system. Transceiver B receives harmonic signal from A is expressed as below:

$$F_{2h} = 2fa, P_{2h} = P - R_{sth2} - L \quad (7)$$

$$F_{3h} = 3fa, P_{3h} = P - R_{sth3} - L \quad (8)$$

Fig. 3 Harmonic interference from transceiver A to B



3.4 Transmitter Intermodulation Interference

If two or more transmitters work at the same car, signals transmitted from one transmitter will be received by another, and transmitter intermodulation (TXIM) interference will be generated in the last one [9].

The TXIM interference happens as:

1. Signals transmitted from A and B are received by C. TXIM interference is generated as $2fa - fb$ and $2fb - fa$ in C.
2. Signals transmitted from A is received by B. TXIM interference is generated as $2fa - fb$ and $2fb - fa$ in B as shown in Fig. 4a.

Obviously, the second interference is worse. So, it must be considered in EMCA.

Assume that the intermodulation loss is L_I , then the third-order intermodulation interference caused in transceiver B and received by transceiver C is

$$P_{\text{TXIM2}} = P - (L_{AB} + L_I + L_{BC}). \quad (9)$$

3.5 Receiver Intermodulation Interference

A transceiver receives not only the wanted signal but also unwanted signals. Unwanted signals pass through non-linear components and become receiver intermodulation (RXIM) interference [9].

RXIM interference happens as:

1. D receives unwanted signals transmitted from A and B and generates RXIM interference $2fa \pm fb$, $2fb \pm fa$, as shown in Fig. 4b.
2. D receives unwanted signals transmitted from transceiver A, B, and C and generates RXIM interference $fa + fb - fc$, $fa + fc - fb$, $fb + fc - fa$.

The two-signal third-order RXIM interference power and three-signal third-order RXIM interference power caused in transceiver D are in (10) and (11), respectively [9].

$$P_{\text{RXIM2}} = 2(P - L_{AD}) + (P - L_{BD}) + C_{2,3} - 60 \times \lg(\Delta f) \quad (10)$$

$$P_{\text{RXIM3}} = (P - L_{AD}) + (P - L_{BD}) + (P - L_{CD}) - 81 \times \lg(\Delta f) \quad (11)$$

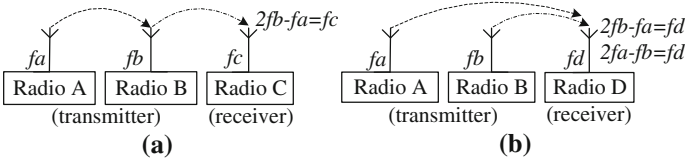


Fig. 4 **a** Transmitter intermodulation interference. **b** Receiver intermodulation interference

$C_{2,3}$ is two-signal third-order intermodulation constant. $\Delta f = (|fd - fa| + |fd - fb|)/2$ (MHz) in (10) and $\Delta f = (|fd - fa| + |fd - fb| + |fd - fc|)/3$ (MHz) in (11).

According to above analyses, co-channel/adjacent channel interference, harmonic interference, and TXIM and RXIM interference must be considered in EMCA. The flowchart of EMCA is shown in Fig. 5.

4 Subnet QoS Model Based on EMCA

4.1 Precise EMC QoS Model

The subnet topology T_i is composed of N clusters. Each cluster uses one frequency at each moment. So the N -clusters topology has $\prod_N |F_j|$ spectrum combinations, defined as $T_{POOL}(\{F_j\})$.

T_{POOL} is the first factor of EMC QoS model. But a topology with big pool is not always better than a topology with small pool. If one frequency in cluster is not compatible electromagnetically with any frequency in other clusters, the usable frequency groups are not as big as before. So the second factor of QoS model is introduced to solve this problem. This factor is denoted by C_{POOL} , which is a subset of T_{POOL} satisfying EMC.

A subnet has bottleneck if one of the clusters has very little frequencies. To avoid the subnet failed by the failure of such clusters, the third factor of QoS model is introduced as $L_{POOL}(\{F_j\})$, which is the minimal number of a subset of C_{POOL} in terms of one specified cluster.

Considering the above three factors, we get the pEMC QoS model in (12).

$$T_i \succ T'_i \Leftrightarrow$$

$$L_{POOL}(\{F_i\}) > L_{POOL}(\{F'_i\}) \vee \left(\left(\begin{array}{l} L_{POOL}(\{F_i\}) = L_{POOL}(\{F'_i\}) \wedge \\ \left(C_{POOL}(\{F_i\}) > C_{POOL}(\{F'_i\}) \vee \right. \\ \left. \left(C_{POOL}(\{F_i\}) = C_{POOL}(\{F'_i\}) \wedge \right. \\ \left. \left. T_{POOL}(\{F_i\}) < T_{POOL}(\{F'_i\}) \right) \right) \right) \right) \quad (12)$$

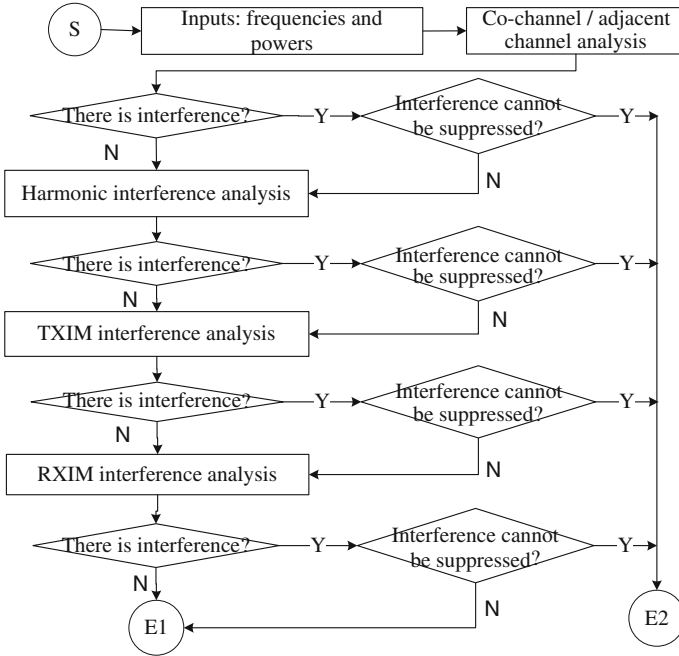


Fig. 5 The flowchart of EMCA

4.2 Approximate EMC QoS Model

As the complexity of EMC analysis in pEMC model is tight related with the number of usable frequency between *PSU* and *CSUs*, a real-time spectrum decision based on pEMC model cannot be guaranteed when there is a large usable frequency set. Then, the aEMC model is proposed. The approximate model does not consider the electromagnetic compatibility, as seen in (13).

$$T_i \succ T'_i \Leftrightarrow L_{\text{POOL}}(\{F_i\}) > L_{\text{POOL}}(\{F'_i\}) \vee \left(L_{\text{POOL}}(\{F_i\}) = L_{\text{POOL}}(\{F'_i\}) \wedge T_{\text{POOL}}(\{F_i\}) \geq T_{\text{POOL}}(\{F'_i\}) \right) \tag{13}$$

If the probability of EMC is high for any combination of frequencies, the aEMC model has feasibility. From a theoretic proof, we achieve $P_{\text{EMC}} \approx 59\%$. From the result of simulation 1 in Sect. 6, we get $P_{\text{EMC}} \approx 62.2\%$. The error of two results is smaller than 5% and is acceptable.

5 Simulation Design

Before simulation, we assume the below scene. There is a $(1 + N)$ -user subnet, which means there are 1 *PSU* and N *CSUs*. Each user has N transceivers and 1 common control channel (CCC) transceiver which uses 1–5 frequencies. The same transceiver has the same performance. We design three simulations in this part.

5.1 Simulation 1

In simulation 1, frequencies are chosen randomly to simulate the probability of EMC.

Description:

```

1: for 1..loop1
2:   Set parameters of CCC and other transceivers
3:   for 1..loop2
4:     choose PSU's transceiver randomly
5:     assign transceivers' parameters randomly
6:     if incompatible then  $cnt = cnt + 1$ 
7:   return  $P_{EMC} = 1 - cnt / (loop1 * loop2)$ 

```

5.2 Simulation 2

In simulation 2, the coherence between aEMC model and pEMC model is simulated for any two topologies. First two topologies are generated randomly, and then, the QoS values based on two models are compared and denoted as T (true) or F (false). This process is repeated for m times and the number m' of result with T is counted. The reliability of aEMC model is expressed by $r_{aEMC} = m'/m$.

Description:

```

1: for 1..loop1
2:   Set parameters of CCC and other transceivers
3:   generate  $N$  randomly to determine the subnet size
4:   generate common frequency set  $Fb_k$  of PSU and CSUs
5:   generate clusters' spectrum pool  $F_j$  based on  $Fb_k$ 
6:   for 1..loop2
7:     generate topology  $T_A$  and topology  $T_B$  randomly
8:     calculate  $aEMC(T_A, T_B)$  ;
9:     calculate  $pEMC(T_A, T_B)$  ;
10:    if  $aEMC(T_A, T_B) = pEMC(T_A, T_B)$  then  $m' = m' + 1$ ;
11:     $m = m + 1$ ;
12:   return  $m'/m$ 

```

5.3 Simulation 3

In simulation 3, the coherence between aEMC model and pEMC model is simulated for all topologies.

A subnet has Nt types of topologies and each type has nt topologies. For each type, we firstly find the best topology T_{BEST} using aEMC model, then compare the others with T_{BEST} based on pEMC, and denoted as T (true) or F (false). The number m' of results with T , and the number of all topology as m are counted. This process is repeated for n times and the reliability of aEMC is expressed as $r_{aEMC} = 1 - m'/m/n$.

Description:

- 1: Set subnet size is $1+N$
- 2: **for** $1..n$
- 3: Set parameters of CCC and other transceivers
- 4: generate Fb_k of PSU and $CSUs$
- 5: generate F_j based on Fb_k
- 6: calculate Nt based on N
- 7: **for** $1..Nt$
- 8: calculate T_{BEST} using aEMC model
- 9: **for** $1..nt-1$
- 10: **if** there is other topology better than T_{BEST} using pEMC model **then**
- 11: $m' = m' + 1$;
- 12: $m = m + nt - 1$
- 13: **return** $1 - m' / m$;

5.4 Explain for Simulation Parameters

The expression of a topology affects the generation method directly. The topologies which have the same number of $CSUs$ are the same type of topology. In each type of topologies, $CSUs$ are permuted and combined. To avoid repetition of topology between different types, the below rules are followed:

1. The $(i + 1)$ th cluster cannot smaller than i th cluster.
2. Sort the possible $CSUs$ combination in clusters. The position of CSU combination in $(i + 1)$ th cluster cannot be in front of the position of CSU combination in i th cluster.

The position of CSU combination in a cluster is expressed by a 0–1 matrix, in which 1 expresses that the CSU in this position is in the cluster. Take a 7-users subnet as example, the combinations of 2- $CSUs$ clusters can be expressed as below. To ensure the public frequency set is not empty, $|Fb_k|$ is chosen large enough.

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 1 & 0 & 1 & 0 & 0 \end{pmatrix}$$

6 Simulation Results

6.1 Result of Simulation 1

In simulation 1, the CSUs' number is chosen randomly between 1 to 6. As in Table 1, the probability of compatible electromagnetically is $P_{\text{EMC}} \approx 62.2\%$, showing that the pEMC model is feasible. The running time of pEMC model in a 5-users subnet, if $|Fb_k| = 1,000\text{--}1,500$, needs almost 400 days to complete the EMCA.

6.2 Result of Simulation 2

In simulation 2, only the subnets whose users' number is less than five are considered.

According to Table 2, we can get $r_{\text{aEMC}} = 97.47\%$. Therefore, there is a high consistency between aEMC model and pEMC model.

6.3 Result of Simulation 3

Based on the conclusion in 6.1, the subnets whose size is less than four are considered in this simulation.

According to Table 3, in a 3-users subnet, $r_{\text{aEMC}} = 97.2\%$, and in a 4-users subnet, $r_{\text{aEMC}} = 96.3\%$. The results are close to the result in simulation 2.

Table 1 Data of simulation 1

loop1	loop2	cnt	Time (s)
1,000	1,000	379,852	7
1,000	1,000	379,102	7
1,000	1,000	376,130	7
10,000	10,000	37,771,785	687
10,000	10,000	37,765,125	686
10,000	10,000	37,794,096	686

Table 2 Data of simulation 2

loop1	N	loop2	Each m'
10	2	8	8
	3	20	20
	3	20	19
	4	30	30
	4	30	28
	3	20	20
	1	1	1
	2	8	8
	1	1	1
	3	20	19

Table 3 Data of simulation 3

N	n	m'	m
2	1,000	29	1,000
3	1	2	47
4	10	17	470

7 Conclusion

The spectrum decision in CR must consider the QoS requirement. QoS in this paper is to ensure the largest network and the optimal parameter for communication. The principle and process of EMCA are proposed first. Then, the pEMC model and aEMC model are built, respectively. At last, the consistency of two models is simulated. The result of simulations shows that pEMC model can be replaced by aEMC model.

Acknowledgments This research is supported by the State Key Laboratory of Integrated Information System Technology, Institute of Software, Chinese Academy of Sciences. The author wishes to thank Zhou Xin, an engineer in the laboratory, who presented related application requirements; Doctor Liao Ming-Xue, who proposed the EMC QoS models and related simulation designs; and Xu Kun, a graduate student, who completed a theoretic proof.

References

1. Mitola J et al (1999) Cognitive radio: making software radios more personal. *IEEE Pers Commun* 6(4):13–18
2. Mitola J (2000) Cognitive radio: an integrated agent architecture for software defined radio. Royal Institute of Technology (KTH)
3. Federal Communications Commission (2002) Spectrum policy task force. Rep ET Docket no. 02–135, Nov 2002
4. Haykin S (2005) Cognitive radio: brain-empowered wireless communications. *IEEE J Sel Areas Commun* 23:201–220

5. Liao MX, He XX, Jiang XH (2012) Optimal algorithm for cognitive spectrum decision making. In COCORA 2012, pp 50–56
6. Ji PP, Liao MX, He XX, Deng Y (2011) Extreme maximal weighted frequent itemset mining for cognitive spectrum decision making. In: IEEE ICCSNT 2011, pp 267–271
7. Fan ZJ, Liao MX, He XX, Hu HH, Zhou X (2011) Efficient algorithm for extreme maximal biclique mining in cognitive spectrum decision making. In: IEEE ICCSN 2011, pp 25–30
8. McMahan JH (1974) Interference and propagation formulas and tables used in the Federal Communications Commission spectrum management task force land mobile frequency assignment model. *IEEE Trans Veh Technol* 23(4):129–134
9. Hanna SA (1998) Intermodulation and harmonic analysis for land mobile radios communications. In: *IEEE VTC 1998*, vol 1. pp 288–292

Separating and Recognizing Gestural Strokes for Sketch-Based Interfaces

Yougen Zhang, Hanchen Song, Wei Deng and Lingda Wu

Abstract Gestures are widely used as shortcuts to invoke commands in pen-enabled systems. The pen mode problem needs to be addressed when integrating gestures into sketching applications. Traditionally, explicit mode switching methods are widely employed to separate gestural strokes from normal inking strokes. However, explicit methods may have availability constraints or be inefficient under many circumstances. In this work, a new gesture interaction paradigm for sketch-based interfaces was proposed. The lasso stroke, which is widely used for selection, was modified and used to indicate the entrance of gesture mode. Our approach is intuitive and efficient; it could be a useful addition to pen-based user interfaces.

Keywords Sketch-based interface · Pen gesture · Mode switching · Gesture recognition

1 Introduction

Pen-based user interface is a primary kind of post-WIMP (window icon menu pointer) interface. It allows for fluid and expressive input based on the pen-and-paper metaphor in tasks such as design sketching, note taking, and computer

Y. Zhang (✉) · H. Song
Science and Technology on Information Systems Engineering Laboratory,
National University of Defense Technology, 410073 Changsha, China
e-mail: zhangyougen@nudt.edu.cn

H. Song
e-mail: songhanchen@hotmail.com

W. Deng · L. Wu
Key Laboratory, Academy of Equipment, 101416 Beijing, China
e-mail: dengw_021@163.com

L. Wu
e-mail: wulingda@139.com

operating. In recent years, with the fast development and growing popularity of electronic whiteboard, smart phones, PDAs, and Tablet PCs, pen-based interfaces are becoming more and more prevalent [1]. Gesturing is an important means of interaction in pen-enabled interfaces. The gesture strokes are widely used as shortcuts to invoke commands. Compared with traditional GUI elements (menu, toolbar button, and keyboard shortcut), pen gestures are efficient yet cheap, since they require little additional hardware resources, like screen space or hardware buttons.

This work aims at integrating pen gestures into sketching applications. There are several issues that need to be addressed. Firstly, a set of gestures should be designed according to the desired functionalities, which is specific to application. Secondly, the gesture recognizer that distinguish input gestural strokes need to be constructed. This problem has been extensively studied, and relatively robust solutions are widely available [2]. Another important issue is the so-called stroke mode problem.

Since both sketching and gesturing are done stroke by stroke with the pen, input strokes could be either ink strokes or gestural strokes. The former are data that should be stored for later processing; the latter are commands that intended for immediate interpretation and execution by the computer [2]. Therefore, it is essential to determine whether the system is in ink mode or in gesture mode for each input stroke, so as to decide how this stroke should be processed. For example, a design-by-sketch system allows users to draw sketch strokes naturally and to edit the sketch by issuing editing gestures (copy, paste, delete, etc.). In the designing process, users may switch between ink mode and gesture mode frequently and irregularly. Obviously, an ineffectual ink/gesture mode switching technique may become the bottleneck in the system usability.

The mode problem faced by these systems is a classic problem [3]. It has long been considered as an important source of errors, confusion, unnecessary restrictions, and complexity [4]. Traditional solutions to this problem simply require the user to switch modes explicitly. For example, a toolbar with icons may be employed, on which the user taps to enter the intended mode. However, the resulting round-trip time interrupts the user's attention from his/her work [5]. Pen barrel button and tablet bezel button are also often used for explicit mode switching [4] in many existing pen-based applications, usually pressing button for gesturing. However, these supplementary physical buttons are not always available. Even if a button is available, mode errors occur if the user forgets to press the button prior to inputting his/her sketch or command stroke, because the mode switching action is mentally separated with the gesture actions. That will result in a spurious ink stroke or an unexpected command recognized and executed, depending on the error type. To recover from the error, the user has to disrupt his/her task, check and modify the digital ink content, switch to the right mode, and then repeat the intended input stroke [3].

In this paper, we proposed a solution of gesture interaction for sketching interfaces, aiming at reducing the burden of mode switching on users. The main idea of our design is to use a detectable lasso stroke for indicating the entrance of

gesture mode in addition to target selection. The remainder of this paper is organized as follows: [Sect. 2](#) briefly introduces the previous research on ink/gesture mode switching problem. In [Sect. 3](#), we present the design of our generally applicable gesture interface; its key design concepts are discussed. [Section 4](#) describes some issues on recognition. In [Sect. 5](#), we conduct a preliminary evaluation. Conclusion and future study are discussed in [Sect. 6](#).

2 Related Work

Various methods have been investigated for performing stroke mode switching. As discussed in the previous section, explicit mode switching methods are simple and widely employed, but they may have availability constraints or be inefficient under many circumstances. Therefore, efforts have also been made to the research of implicit mode switching techniques in recent years.

Implicit mode switching aims at releasing user's physical and mental burden of switching between modes manually. Strictly speaking, without knowing the mental state of the user, the problem of implicitly distinguishing strokes intended to be gestures from those intended to be ink is not computable; however, by making assumptions about the likelihood of certain interaction sequences, it is possible to build actually effective systems [6].

Nijboer et al. [7] explored using frame gestures to control rotation, translation, and scale of the drawing canvas and of stroke selections. By mapping canonical transformations (translation, rotation, scaling) of the canvas or of stroke selections to contextual gestures that are started from the canvas border or the selection frame, frame gestures enable a fluid switching between normal drawing, interaction with the drawn strokes, and interaction with the canvas, without having to switch between dedicated operating modes. A limitation of this design is the restriction by the actual interface border. Furthermore, frame gestures are just applicable to transformations but not other gesture commands.

Li et al. [4] explored using pen pressure that is available on many tablet devices to achieve an implicit mode switching. They leave the heavy spectrum of the pressure space for gesturing and preserve the normal (middle) pressure space for inking. Since users have differences in their inherent pressure spaces, personalized pressure spaces are needed. Analogously, 3D orientation of the pen was also studied to address the mode problem [8].

Saund and Lank [3] present a solution to the mode problem in pen-based programs. They offered an inferred-mode interaction protocol that avoids the prior selection of mode during inking. The system tried to infer the user's intent, if possible, from the properties and context of the pen trajectory. When the intent is ambiguous, a choice mediator for the user is offered, which can also be ignored so as to maximize the fluidity of drawing. A similar design called "Handle Flags" was proposed by Grossman et al. [5]. When the user positions the pen near an ink stroke, Handle Flags are displayed for the user to perform potential selections.

However, their techniques only avoid prior mode selection for the selection sub-task. Another limitation of Handle Flags is that they rely heavily on the result of the stroke grouping algorithm, which can be complicated in unstructured diagrams.

Zeleznik and Miller proposed a design, which is called “Fluid Inking” [6], for disambiguating gestures from regular inking. This design uses two simple patterns: the prefix flicks (fast straight lines) and post-fix terminal punctuation (fast taps or short pauses). We also use a lasso stroke as gesture prefix, but it is integrated with the selection operation, and it is consistent among all gestures.

Guo and Chen investigated the recognition of handwriting-editing gestures and alphanumeric in a mixed recognition mode [9]. Gestures are mixed directly with handwriting strokes. As a result, the recognition system is prone to be confused.

3 Our Design of Gesture Interaction

Our goal is to develop a gesture interface for a sketch-based military situation marking system. As discussed in previous sections, the essence of the stroke mode problem is a compromise between the freedom of user operation and the complexity of mode inference/recognition for the system. In order to provide users with an approach of switching stroke modes quickly and conveniently, we focus on the design of fluid mode transitions according to sketching interaction patterns.

In this section, a simple analysis of gestures in the sketching interface is provided. We propose to use the lasso stroke, which is widely used as the selection gesture, to indicate the entrance of gesture mode. The new gesture issuing method is introduced.

3.1 Gesture Set Analysis

Empirical analysis shows that, besides normal inking of text and graphics, three major types of operations are desired and frequently performed in the process of military situation marking: (1) editing objects, including copy, paste, delete, move, rotate, resize, or label objects. Object here may be either raw strokes or recognized text/symbols; (2) manipulating the background map, such as zoom in/out or translate the map. Map is often used as the reference in situation plotting; (3) other commands provided by the system, for example, save, redo, undo, last/next view, and so on.

We design a set of pen gestures for the sketch-based military situation marking system, as listed in Fig. 1. The red dot of each stroke is the starting point of the gesture trajectory.

Selection is one of the most elementary operations in gesture interactions. Most command involves a selection task explicitly or implicitly. For example, the target object should be specified when performing editing operations such as copy,

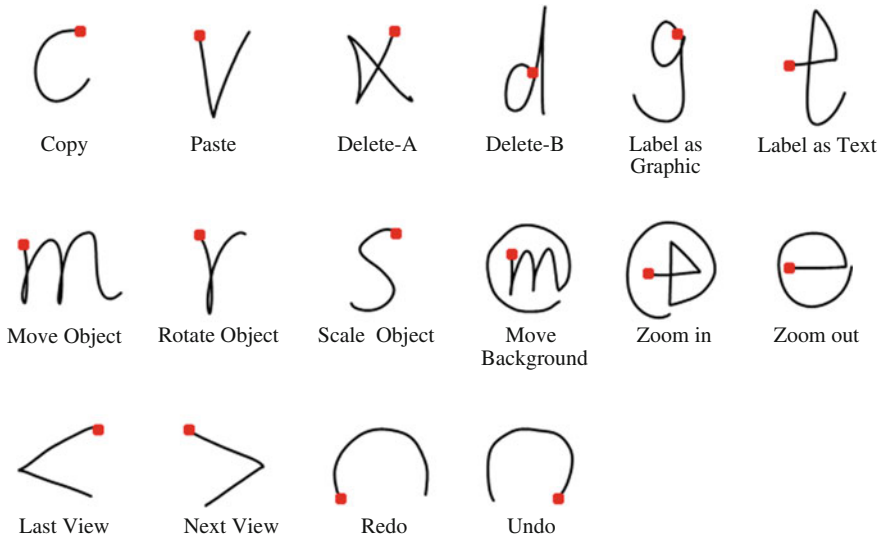


Fig. 1 Gestures designed for the sketching system

delete, etc. When pasting an object, the target location should be specified. As long as most gestures begin with selection, it guided us to the idea of entering gesture mode by detecting selections.

3.2 Gesture Issuing Method

Lassoing is widely used for selection both in pen-based interfaces and in our daily life. It allows users to select by drawing an enclosing stroke around target objects. Lassoing works well especially for small scattered targets [10], which is often the case in sketched diagrams that contain numerous short strokes.

We hereby set several conventions for issuing gestures. By default, the sketching system works in ink mode. When a gesture task is required, users can take the following steps: Firstly, draw a lasso stroke and enclose target objects or location if necessary. Secondly, draw the gesture command stroke inside the lasso. This stroke would be fed to the gesture recognizer which outputs the class label of the gesture. For most gestures, the corresponding command would be executed then. Besides instantly executable commands, there are also some gestures suitable for interactive operation. These mainly include object transformation gestures (move, resize, rotate) and background manipulation gestures (move, zoom in, zoom out). In this case, users should draw an additional stroke to specify further parameters of the command. For example, when moving an object, this stroke is used to drag the object to somewhere interactively. Analogously, the degree of

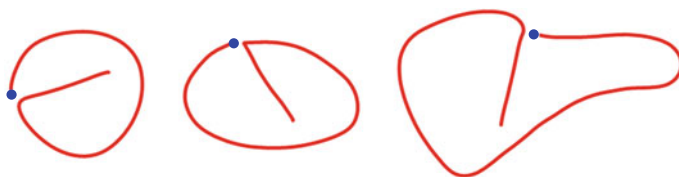


Fig. 2 Examples of lasso stroke

other object transformation gestures and background manipulation gestures could also be determined in this way.

Based on the above interaction conventions, the system can be aware of the stroke mode transitions. Once a lasso stroke is detected, the system enters gesture mode from normal ink mode and parses the following one stroke as a gesture command. Then it executes the command either instantly or interactively in response to the additional stroke. After that, the system returns to ink mode, waiting for new ink strokes or gesture interaction sessions.

So far, the problem of detecting gesture strokes comes down to detection of the lasso stroke. The usual lasso stroke we used is simply a closed stroke, which can be easily confused with ordinary circle in ink of drawing. Therefore, we use a modified lasso stroke that is reliably distinguishable from normal inking strokes. As shown in Fig. 2, our lasso stroke is extended with a straight segment inside the original closed stroke. This minor modification eliminates most confusion. The lasso stroke can be drawn either clockwise or counterclockwise, of deformable shape according to the shape/distribution of the target object(s). The next section will discuss the detection of the lasso stroke, as well as the recognition of gesture command strokes.

4 Recognition

Gesture recognition is the process of parsing a hand-drawn stroke as being one of the predefined gesture types. According to our gesture issuing method described in Sect. 3, a gesture consists of one lasso stroke and one or two command strokes.

4.1 Lasso Stroke Recognition

The lasso stroke acts as the sign of entering gesture mode and also specifies the target of the gesture in most cases. Therefore, it is critical to detect the lasso stroke in the sequence of input strokes quickly and accurately. In order to accommodate target object(s) and user habits, the lasso stroke is allowed to be deformable and be either clockwise or counterclockwise. So it is not feasible to detect the lasso stroke

by matching it to predefined templates. We treat this as a binary classification problem and used a feature-based method to test whether an input stroke s is a lasso stroke.

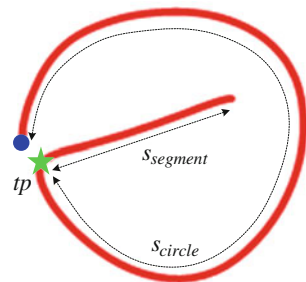
Several features of s were extracted as follows:

- SL Stroke length of s
- DE Distance between end points of s
- LB Diagonal length of the bounding box of s
- MR Minimum of the ratio of distance/length. Here, both the distance and the length are measured from a sample point to the start point. When s is a lasso stroke, MR locates the turning point between the sub-stroke of circle and the sub-stroke of straight segment. Thus, we denote the point that takes MR by tp ; denote the sub-strokes before and after tp by s_{circle} and $s_{segment}$, respectively.
- PI Percentage of $s_{segment}$ that lie inside the bounding box of s_{circle}
- LT Location of tp , that is, the index of tp among sample points, normalized to the range of $[0,1]$.
- SS Straightness of $s_{segment}$, defined as the ratio of the distance between the end points to the length of $s_{segment}$

We collected a set of training strokes. It consists of lasso strokes under different circumstances, normal inking strokes, and inking strokes that may confuse with the lasso stroke (for example, circle, symbol θ and φ). Then a decision tree classifier was trained. This classifier works well with precision and recalls of 99.4 and 97.2 %, respectively, in 10-fold cross-validation, consuming less than one millisecond per stroke (Fig. 3).

Once an input stroke was classified as lasso stroke, it would be highlighted in different color and stroke width. This provides users with instant feedback on the state of stroke mode switching, allowing users to recover immediately from possible false rejection or false acceptance of lasso stroke classification. The lasso stroke is generally distinguishable from normal inking strokes. In rare cases when a lasso-like inking stroke needs to be drawn, users can draw a check “√” on that stroke (yet highlighted) immediately after it was drawn (within the time-out period). Then it would be stored as an inking stroke and displayed in normal color.

Fig. 3 Illustration of feature extraction



For a lasso stroke, we compute its bounding box and convex hull. Then the enclosed objects would be identified and highlighted, indicating the scope of selection. The bounding box and convex hull would also be used later to determine the size and location of the gesture, as well as the relative position between the lasso stroke and its succeeding command stroke.

4.2 Command Stroke Recognition

Since the work of Rubine, numerous pen gesture recognition methods have been proposed, falling into one of two main categories: template based and feature based. Some of the famous methods are Dollar 1, Protractor, and 1¢. The performance of existing methods is reasonably acceptable; therefore, we did not focus on the recognition of unistroke gesture commands. The method we adopted is similar to Dollar 1 developed by Wobbrock et al. except for some modifications in scaling and rotation.

4.3 Gesture Execution

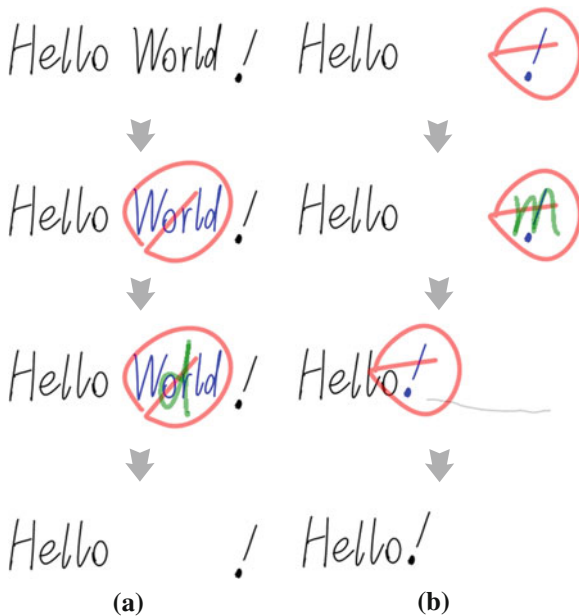
When defining a gesture, each command stroke is mapped to a command. Once the command stroke is recognized, and the corresponding command could be executed immediately in case of instant gesture. If the command stroke corresponds to an interactive gesture, the system waits for the additional stroke, which drives the gesture execution interactively, providing users with feedback of the operation result. The way of parsing the additional stroke is gesture specific.

Figure 4 gives two examples of gesture interactions using the proposed method. The first one is a delete gesture containing a lasso stroke and the command stroke. It is executed directly. Figure 4b illustrates an interactive gesture. When the command stroke is recognized as the “Move,” the system moves the selected strokes following the pen tip in real time when the succeeding stroke is drawn.

5 Evaluation

A preliminary user study was conducted to evaluate the performance of our proposed gesture interface and gain user feedback. We implemented a prototype sketch-based military situation marking system with our gesture interface. A traditional gesture interface was also implemented and tested for comparison. It employs two toolbar buttons for explicit mode switching, and it uses the common circling stroke for object selection. A set of 16 gesture commands (as listed in Fig. 1) were tested. However, several of them were not executable. In other words,

Fig. 4 Two examples of gesture interactions. **a** Delete. **b** Move



the recognition result of these commands were displayed but not executed, because the implementation of them is rather tedious and beyond the scope of this paper.

Six participants took part in the evaluation; all of them were experienced in computer operation, but had little or no experience of using pen-enabled computers except for smart phones. A Lenovo ThinkPad X200 tablet was used as sketching input device. Participants were instructed about the use of our sketch-based marking system on the tablet for several minutes. Then they were asked to perform two gesturing tasks. The first task is to test the supported gestures separately on a given sketched diagram. In the second task, participants were asked to sketch a diagram under textual guidance on operations of inking (draw graphics and write text labels) and gesturing. Upon completion of the tasks, participants were encouraged to make comments and suggestions to obtain additional feedback.

We observed a total of about 400 gesture interactions using each method. When using the proposed method, twelve lasso strokes were misclassified as ink strokes (false negatives). They were mainly careless lasso strokes when users had not yet adapted. There were also three misclassified ink strokes (false positives); they were all corrected automatically since their subsequent strokes were drawn outside of them, thus no accidental gesture was triggered. As for the traditional method, seven mode switching operations were neglected. Thirty-eight gestures were recognized incorrectly, mainly due to errors of our simplified command stroke recognition algorithm, which adopts nearest neighbor matching.

According to the comments and suggestions, participants stated that the proposed gesture issuing method is indeed more complex, but it is still quite intuitive and very easy to memorize and use. Nevertheless, the proposed approach is more efficient because it saves the round-trip time of mode switching. Moreover, the traditional method is error prone since mode switching may be neglected, while in the proposed approach mode, transitions are automatic, allowing users to concentrate on inking and gesturing.

The user study also helped us in identifying some limitations and possible improvements of our approach. A limitation of our lasso-based selection is its poor performance under extreme conditions. For example, it is difficult to select individual objects using a lasso in densely crowded conditions. Furthermore, one participant stated that it is inefficient to draw a lasso stroke that encloses a large-size object (like a long curving stroke). In these cases, the tapping-based selection method is more desirable. We consider adding a select command to our gesturing framework, so as to support more flexible selection. Moreover, we can identify a large-size object as selected if it is covered by the lasso stroke in its center over a certain area. Another suggested improvement is to allow for manual explicit mode switching as a complement to the proposed approach, so as to address the case of consecutive gesture interactions.

6 Conclusions and Future Work

We have presented a novel gesturing approach for sketch-based interfaces. The main advantage over traditional explicit mode switching approaches is that it supports fluid mode switching without extra hardware requirements. This makes it applicable to various pen-enabled systems ranging from tablets and smart phones to electronic whiteboards. Moreover, our gesture issuing method is consistent among gestures, thus lighten the burden of memory load compared to the method of [6]. The preliminary evaluation results indicate that participants learned to use our interface in a short period of time.

Our future work will focus on the personalization of command strokes. Since we have used template-based approach for recognizing these unistroke gestures, it is straightforward to define personalized gestures by storing user-specific templates. However, some issues remain to be studied, for example, how to help user choosing gestures so as to avoid possible conflict with existing ones and how to organize templates efficiently. Furthermore, the presented user study was informal so far, and further study is necessary in order to gain more useful feedback.

Acknowledgments This work is supported by the National Science Foundation of China under Grant No. 61103081.

References

1. Johnson G, Gross MD, Hong J, Do EY-L (2009) Computational support for sketching in design: a review. *Found Trends Human Comput Interact* 2:1–93
2. Appert C, Zhai S (2009) Using strokes as command shortcuts: cognitive benefits and toolkit support. 27th international conference on Human factors in computing systems. ACM, Boston, pp 2289–2298
3. Saund E, Lank E (2003) Stylus input and editing without prior selection of mode. 16th annual ACM symposium on user interface software and technology, Vancouver, Canada. ACM, pp 213–216
4. Li Y, Hinckley K, Guan Z, Landay JA (2005) Experimental analysis of mode switching techniques in pen-based user interfaces. SIGCHI conference on human factors in computing systems, Portland, OR, USA, pp 461–470
5. Grossman T, Baudisch P, Hinckley K (2009) Handle flags: efficient and flexible selections for inking applications. *Graphics interface 2009*, Canadian information processing society, Kelowna, British Columbia, Canada, pp 167–174
6. Zeleznik R, Miller T (2006) Fluid inking: augmenting the medium of free-form inking with gestures. *Graphics interface 2006*. Canadian information processing society, Quebec, Canada, pp 155–162
7. Nijboer M, Gerl M, Isenberg T (2010) Exploring frame gestures for fluid freehand sketching. Seventh sketch-based interfaces and modeling symposium. Eurographics Association, Annecy, France, pp. 57–62
8. Tian F, Jiang Y, Dai G, Wang H (2009) Instruction method based on stroke tail gesture. In: CAS, I.o.S. CN 200910080176
9. Guo F, Chen S (2010) Gesture recognition techniques in handwriting recognition application. 12th international conference on frontiers in handwriting recognition, Kolkata, pp 142–147
10. Mizobuchi S, Yasumura M (2004) Tapping vs. circling selections on pen-based devices: evidence for different performance-shaping factors. In: Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, Vienna, Austria, pp 607–614

Modeling and Performance Analysis of Workflow Based on Advanced Fuzzy Timing Petri Nets

Huifang Li and Xinfang Cui

Abstract Time management plays an important role in workflow management systems. Aiming at describing all the uncertain time information existing in practical business process, this paper applied fuzzy sets theory to workflow time modeling and performance analysis and then presented an improved fuzzy timing workflow nets. Based on the reduction analysis of fuzzy time performance for several constructions of workflow models, a fuzzy time analysis and reasoning method was proposed and then a hierarchical algorithm for business process temporal reachable probability analysis was proposed. Finally, a case study illustrates that our method is feasible and efficient.

Keywords Fuzzy time constraint · Workflow · Possibility theory · Advanced fuzzy timing workflow nets

1 Introduction

Workflow modeling and model performance analysis are important research contents of workflow. In actual business processes, there are many time constrains, and if violated, it may bring loss to the enterprise, so the timing description ability of workflow models becomes the focus of current workflow modeling studies [1].

Aalst first applied Petri net, which has intuitive graphical representation and solid mathematical foundation, to workflow management and proposed workflow net (WF-net) [2]. Recently, in order to describe and analyze time behaviors in

H. Li (✉) · X. Cui

Automation School, Beijing Institute of Technology University, Beijing, China
e-mail: huifang@bit.edu.cn

X. Cui

e-mail: cuixinfangxiaoxin@163.com

workflow, many Petri net-based timing workflow models are proposed, such as time workflow net(TWF-net) by Ling and Schmidt [3], and timing constraint workflow nets (TCWF-net) by Li and Fan [4]. In these models, time information is certain; however, resources and activities have dynamic characteristics in actual workflow because of the uncertain factors, such as machine fault, different proficiency of workers and so on, so there is a lot of uncertain time information which is hard to describe and analyze. Aiming at this requirement, Murata first applied fuzzy theory to time Petri nets and proposed Fuzzy-timing High-level Petri nets (FTHN) [5]. Pan and Tang [6] added certain effective time constraints on resources and activities and proposed fuzzy temporal workflow nets (FTWF-nets). But in order to improve the flexibility and stability of the dynamic workflow management of real business processes, the effective time constraints attached to resources and activities should have fuzzy values as well. So based on FTWF-nets, by applying high-level fuzzy timed Petri net (HFTN) [7] to workflow modeling, the advanced fuzzy timing workflow nets (AFTWF-nets) model was proposed in this paper and a stratification algorithm of temporal reachable probability of AFTWF-nets model was given.

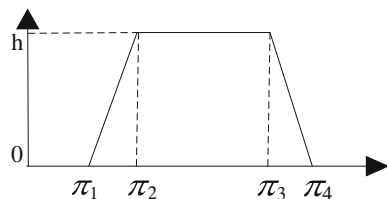
In AFTWF-nets model, fuzzy time was used to describe the effective time constraints of resources and activities, calculate, and analyze. Meanwhile, the stratification algorithm of temporal reachable probability based on AFTWF-nets reduced the complexity of the data structure and made it much easier to develop and realize the workflow management system.

2 Advanced Fuzzy Timing Workflow Nets

2.1 Fuzzy Time Concept

Definition 1 Fuzzy time point is the possibility distribution of a function mapping from the time scale Γ to real interval $[0, 1]$, which restricts the possible value of a time point. Let π_a denotes the possibility function attached to a time point a , then $\forall \tau \in \Gamma, \pi_a(\tau)$ denotes the numerical estimate of the possibility that a is precisely τ . Let fuzzy set A be the possible range of a , and μ_A denote the membership function of A and then we have $\forall \tau \in \Gamma, \pi_a(\tau) = \mu_A(\tau)$. In this paper, a fuzzy time point is denoted by trapezoid possible distribution [6], which must be normal and convex. So a fuzzy time point can be represented by $h[\pi_1, \pi_2, \pi_3, \pi_4], (\pi_1 \leq \pi_2 \leq \pi_3 \leq \pi_4)$ (Fig. 1 shows an example). And it becomes a fixed-length time when $\pi_1 = \pi_2, \pi_3 = \pi_4$ and a fixed time point when $\pi_1 = \pi_2 = \pi_3 = \pi_4$.

Fig. 1 Trapezoid function of fuzzy time point



2.2 AFTWF-nets Model

Definition 2 AFTWF-nets is a 8-tuple $(P, T, F, M_0, FT, FD, FTPC, FTTC)$.

1. (P, T, F, M_0) is a basic workflow net. P is a set of places, T is a set of transitions, $P \cap T = \emptyset, P \cup T \neq \emptyset$. F is a set of arcs which is a subset of $(P \times T) \cup (T \times P)$. M_0 represents the initial marking. Let t denotes a transition, and p denotes a place.
2. FT, fuzzy timestamp, denotes a set of fuzzy timestamps attached to tokens, which represents the possibility distribution of a token's arrival time at a place, and let $\pi(\tau) = h[\pi_1, \pi_2, \pi_3, \pi_4]$ denotes the fuzzy timestamps function.
3. FD, fuzzy delay, denotes a set of fuzzy delays of transitions, which describes the possibility distribution of the duration from t triggering to t outputting tokens to its output places, and let $d_t(\tau) = d[d_1, d_2, d_3, d_4]$ denotes the fuzzy delay function.
4. FTPC, fuzzy token time constraint, denotes a set of valid intervals constraints of tokens, which is signed as FTFC $(p) = 1[a_1, a_2, a_3, a_4]$. Let τ denotes a time point when a token arrives at a place, so the token is definitely valid during time interval $[\tau + a_2, \tau + a_3]$, it is uncertainly valid in $[\tau + a_1, \tau + a_2]$, and it is definitely invalid out of $[\tau + a_1, \tau + a_4]$. If FTFC (p) is a fixed time interval, then $a_1 = a_2$ and $a_3 = a_4$. If there is no time constraint, then FTFC $(p) = 1[0, 0, 0, 0]$.
5. FTTC, fuzzy transition time constraint, denotes a set of valid intervals constraints of transitions, which is signed as FTTC $(t) = 1[b_1, b_2, b_3, b_4]$. Let τ denotes a time point when a transition is enabled, so the transition surely can trigger in $[\tau + b_2, \tau + b_3]$, it possibly triggers in $[\tau + b_3, \tau + b_4]$, and it surely cannot trigger out of $[\tau + b_1, \tau + b_4]$. If FTTC (t) is a fixed time interval, then $b_1 = b_2$ and $b_3 = b_4$. If there is no time constraint, then FTTC $(t) = 1[0, 0, 0, 0]$.

In this model, FD denotes the fuzzy delays of transitions, FTPC limits the life cycle of resources, and FTTC limits the fire time interval of transitions. Such model lays the modeling foundation for the following expanded logical analysis as well as the time-level performance optimization and unification.

3 Performance Analysis of AFTWF-nets

3.1 Time Calculation of AFTWF-nets

Murata. T has given the algorithm of FTN [5], and based on this, AFTWF-nets model redefined and remodeled the time constraint of tokens and transitions, so the new definitions and detailed calculations of time algorithm are shown as follows. In AFTWF-nets, $I_p(t)$ represents input places set of t , $O_p(t)$ represents output places set of t . p_i is the initial place, and p_o is the ending place. In order to facilitate the description and discussion, the workflow net is assumed to be reasonable.

Definition 3 Fuzzy enabled time $e_t(\tau)$: fuzzy enabled time of t denotes the possibility distribution of that t is enabled at time point τ , and $e_t(\tau)$ is decided by the possibility distribution of the latest arrival time of the tokens in $I_p(t)$.

If there is only one token in $I_p(t)$, and it arrives at $\pi(\tau)$, so $e_t(\tau) = \pi(\tau) \oplus \text{FTPC}(p)$

$$\begin{aligned} &= \min\{h, 1\}[\pi_1, \pi_2, \pi_3, \pi_4] \oplus [a_1, a_2, a_3, a_4] \\ &= h[\pi_1 + a_1, \pi_2 + a_2, \pi_3 + a_3, \pi_4 + a_4]. \end{aligned}$$

where: \oplus is extended additive operator [6].

If there are n tokens in $I_p(t)(p_1, p_2 \dots p_n)$, token in p_i arrives at $\pi_i(\tau) = h_i[\pi_{i1}, \pi_{i2}, \pi_{i3}, \pi_{i4}]$ and $\text{FTPC}(p_i) = 1[a_{i1}, a_{i2}, a_{i3}, a_{i4}]$. Then, $e_t(\tau) = \text{latest}\{\pi_i(\tau) \oplus 1[a_{i1}, a_{i2}, a_{i3}, a_{i4}]\} = \min\{h_i\}[\max\{\pi_{i1} + a_{i1}\}, \max\{\pi_{i2} + a_{i2}\}, \max\{\pi_{i3} + a_{i3}\}, \max\{\pi_{i4} + a_{i4}\}], i = 1, 2, \dots n$.

where: latest operator is the trapezoidal function approximation algorithm.

Definition 4 Fuzzy occurrence time $o_t(\tau)$: fuzzy occurrence time of transition t denotes the possibility distribution of that t is fired at time point τ .

If there is no structural conflict, and $\text{FTTC}(t) = 1[b_1, b_2, b_3, b_4]$, then $o_t(\tau) = e_t(\tau) \oplus 1[b_1, b_2, b_3, b_4]$, else there is structural conflicts between m enabled transitions ($t_1, t_2 \dots t_m$), t_i is enabled at $e_{ti}(\tau) = e_i[e_{i1}, e_{i2}, e_{i3}, e_{i4}]$, $\text{FTTC}(t_i) = 1[b_{i1}, b_{i2}, b_{i3}, b_{i4}]$. In order to facilitate discussion, the situation where there is structural conflict between enabled transitions follows “first come first serve” strategy, which means the earlier enabled transition has higher priority. So for t_j , $e_{tj}(\tau) = e_j[e_{j1}, e_{j2}, e_{j3}, e_{j4}]$, then

$$\begin{aligned} o_{tj}(\tau) &= \text{MIN}\{e_{tj}(\tau) \oplus 1[b_{j1}, b_{j2}, b_{j3}, b_{j4}], \text{earliest}\{e_{ti}(\tau) \oplus 1[b_{i1}, b_{i2}, b_{i3}, b_{i4}]\}\} \\ &= \text{MIN}\{e_j[e_{j1}, e_{j2}, e_{j3}, e_{j4}] \oplus 1[b_{j1}, b_{j2}, b_{j3}, b_{j4}], \\ &\quad \text{earliest}\{e_i[e_{i1}, e_{i2}, e_{i3}, e_{i4}] \oplus 1[b_{i1}, b_{i2}, b_{i3}, b_{i4}]\}\} \\ &= \text{MIN}\{e_j[e_{j1} + b_{j1}, e_{j2} + b_{j2}, e_{j3} + b_{j3}, e_{j4} + b_{j4}], \\ &\quad \max\{e_i\}[\min\{e_{i1} + b_{i1}\}, \min\{e_{i2} + b_{i2}\}, \\ &\quad \min\{e_{i3} + b_{i3}\}, \min\{e_{i4} + b_{i4}\}], \}, i = 1, 2, \dots m. \end{aligned}$$

where: the earliest operator means to pick up the earliest enable time from m enable time points and is the trapezoidal function approximation algorithm. MIN is the operator to seek the intersection of possibility distributions.

Definition 5 Fuzzy execution delay $d_t(\tau)$: fuzzy execution delay of transition t is attached on the output arc of t , which denotes the possibility distribution of that it costs τ to finish t . If $o_t(\tau) = o[o_1, o_2, o_3, o_4]$, and $d_t(\tau) = d[d_1, d_2, d_3, d_4]$, then token arrive at $O_p(t)$ at $\pi(\tau) = o_t(\tau) \oplus d_t(\tau) = \min\{o, d\}[o_1 + d_1, o_2 + d_2, o_3 + d_3, o_4 + d_4]$.

3.2 Time Performance Analysis of AFTWF-nets

In AFTWF-nets, possibility theory is applied to describe time information and analyze time performance. Possibility operator and performance index are defined as follows.

Definition 6 Possibility function [6]: Suppose a and b are fuzzy time points, their possibility distributions are represented by trapezoidal function $\pi_a(\tau) = 1[a_1, a_2, a_3, a_4]$, which is signed as ABCD, and $\pi_b(\tau) = 1[b_1, b_2, b_3, b_4]$, which is signed as EFGH, as shown in Fig. 2. Then, the possibility that fuzzy time b is before a (signed as $b \leq a$) can be: Possibility($b \leq a$) = $\frac{\text{Area}([a, b] \cap b)}{\text{Area}(b)} = \frac{\text{Area(EBCH)}}{\text{Area(EFGH)}}$. Especially, if b is a certain time, as shown in Fig. 3, $\pi_b(\tau) = 1[t, t, t, t]$, then Possibility($a \leq b$) = $\frac{\text{Area(AEFD)}}{\text{Area(ABCD)}}$.

Definition 7 Activities execution time constraint satisfaction is the possibility that an activity's execution time meets the expected time, which means the activity ends within the expected time. In AFTWF-nets, assume a token arrives at p_i at $\pi_i(\tau)$ and expect the transition t finishes at $\pi(\tau)$. The token begins from p_i , goes through a series of transitions and finally arrives at $O_p(t)$ at $\pi_a(\tau)$, which can be worked out using the timing algorithm in Sect. 3.1. So the activities execution time constraint satisfaction is Possibility($\pi_a(\tau) \leq \pi(\tau)$).

Definition 8 Time distance between activities constraints satisfaction is the possibility that the time distance between two activities is no less than or no more than the expected time distance. In this paper, it is represented by the difference of

Fig. 2 $a \leq b$ (b is a fuzzy time point)

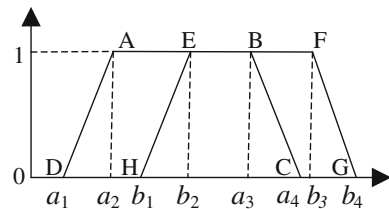
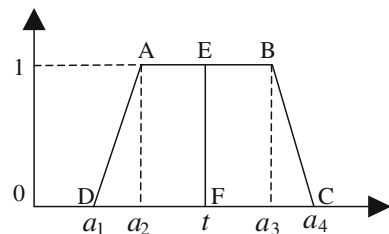


Fig. 3 $a \leq b$ (b is a certain time point)



activities' fuzzy occurrence time. Assume transition t_1 fires before transition t_2 , and the expected time distance is $D(\tau)$. After calculation, we get $o_A(\tau)$ and $o_B(\tau)$, then the possibility that the time distance between t_1 and t_2 is no less than $D(\tau)$ is Possibility($\pi_a(\tau) \leq \pi(\tau)$) and the possibility that the time distance between t_1 and t_2 is no more than $D(\tau)$ is Possibility($o_B(\tau) \leq o_A(\tau) \oplus D(\tau)$).

Definition 9 Temporal reachable probability of process is the possibility that a workflow process finishes within expected time distance. Assume a token arrives at p_i at $\pi_i(\tau)$, and the expected time distance is $D(\tau)$, which means the token is expected to arrive at p_o at $\pi_i(\tau) \oplus D(\tau)$. The token begins from p_i , goes through a series of transitions and finally arrives at p_o , then we can work out $\pi_o(\tau)$. So the temporal reachable probability of process is Possibility($\pi_{p_o}(\tau) \leq \pi_{p_i}(\tau) \oplus D(\tau)$).

3.3 A Stratification Algorithm of Temporal Reachable Probability

In this paper, a stratification algorithm of temporal reachable probability analysis in AFTWF-nets is proposed, which stratifies and extracts AFTWF-nets into hierarchical advanced fuzzy timing workflow nets (HAFTWF-nets) based on the four control route structures of workflow, which are sequence, selecting, parallel, and circle route. Here are the rules of transforming AFTWF-nets to HAFTWF-nets as follow.

1. For sequence route, merge the sequence route without any branch or loop node, represent it in the form of a group of subnets, and represent it by a "place->transition->place" structure named "sequence" in the original net.
2. For parallel route, represent all concurrent branches between "And split" and "And join" in the form of a group of subnets and represent it by a "place->transition->place" structure named "parallel" in the original net.
3. For selecting route, represent all conditional branches between "Or split" and "Or join" in the form of a group of subnets and represent it by a "place->transition->place" structure named "select" in the original net.
4. For circle route, represent the loop in the form of a group of subnets, and represent it by a "place->transition->place" structure named "circle" in the original net.
5. Each route structure can be nested within each other.

Based on the above transforming rules, the concrete steps of stratification algorithm are shown as follows (Fig. 4):

Step 1: Visit p_i .

Step 2: Visit the following nodes successively. When there is parallel, selecting or circle structure, extract the sequence structure before them, then visit their join nodes, do step 3; when visiting p_o , do step 4.

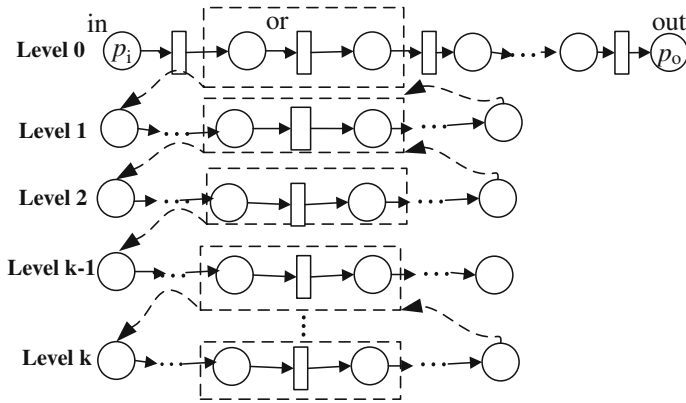


Fig. 4 AFTWF-nets to HAFTWF-nets

Step 3: Extract parallel, selecting and circle structure in subnet form and return to step 2.

Step 4: End visiting. Use the above algorithm in dealing with every subnet successively until there is only sequence structure after simplification.

AFTWF-nets model is called original net, and after transforming, it is called level 0 subnet. Assume the bottom is level k subnet, which contains four basic routes, so when calculating temporal reachable probability, it is needed to calculate level 0 subnet. The simplification and calculation of the basic route structures are shown as follows:

Sequence route: Assume there are n transitions t_1, t_2, \dots, t_n in sequence pattern. So when representing this subnet shown in Fig. 5a as b with p_s, t_s, p_{s+1} , the token in p_{s+1} arrives at: $\pi_{p_{s+1}}(\tau) = \pi_{p_s}(\tau) \oplus \sum_{i=1}^n \oplus (\text{FTPC}(p_i) \oplus \text{FTTC}(t_i) \oplus d_{it}(\tau))$.

where $\sum_{i=1}^n \oplus$ denotes continuous extended plus, $\pi_{p_s}(\tau) = \pi_{p_1}(\tau), i = 1, 2, \dots, n$.

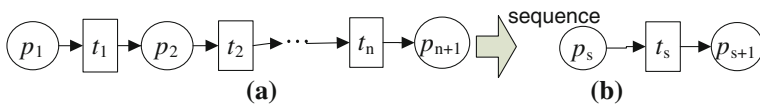


Fig. 5 Simplification of sequence route

Parallel route: Assume there are n transitions t_1, t_2, \dots, t_n in parallel pattern. t_{as} is “And split”, t_{aj} is “And join”, $p_{i1} \in I_p(t_{is})$ and $p_{i2} \in O_p(t_{is})$. So when representing this subnet shown in Fig. 6a as b with p_p, t_p, p_{p+1} , the token in p_{p+1} arrives at:

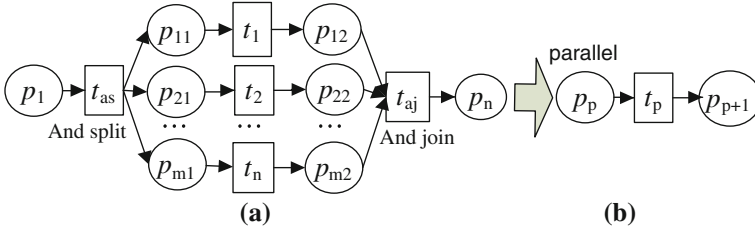


Fig. 6 Simplification of parallel route

$$\pi_{pp+1}(\tau) = \pi_{pp}(\tau) \oplus \text{FTPC}(p_p) \oplus \text{FTTC}(t_{as}) \oplus d_{tas}(\tau) \oplus \text{latest}\left\{\sum_{i=1}^n \oplus (\text{FTPC}(p_{i1}) \oplus \text{FTTC}(t_i) \oplus d_{ti}(\tau) \oplus \text{FTPC}(p_{i2}))\right\} \oplus \text{FTTC}(t_{aj}) \oplus d_{taj}(\tau).$$

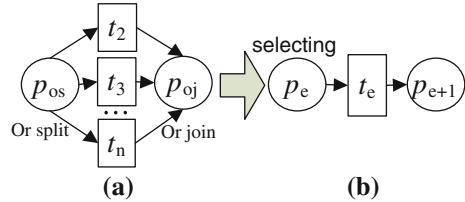
Where $\pi_{pp}(\tau) = \pi_{p1}(\tau)$, $\text{FTPC}(p_p) = \text{FTPC}(p_1)$, $i = 1, 2, \dots, n$.

Selecting route: Assume there are n transitions t_1, t_2, \dots, t_n in selecting pattern. p_{os} is “Or split”, p_{oj} is “Or join”. And the possibility of selecting branch i , which contains t_i , is P_i , where $\sum P_i = 1$. So when representing this subnet shown in Fig. 7a as b with

$$p_e, t_e, p_{e+1}, \text{ the token in } p_{e+1} \text{ arrives at: } \pi_{pe+1}(\tau) = \sum_{i=1}^n (P_i \times (\text{FTPC}(p_e) \oplus \text{FTTC}(t_i) \oplus d_{ti}(\tau)))$$

where $\pi_{pe}(\tau) = \pi_{pos}(\tau)$, $\text{FTPC}(p_e) = \text{FTPC}(p_{os})$, $i = 1, 2, \dots, n$.

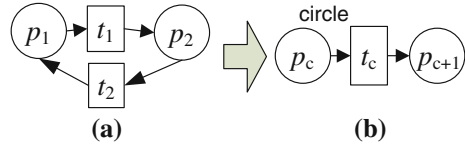
Fig. 7 Simplification of selecting route



Circle route: Assume there are two transitions t_1, t_2 , in circle pattern it is P_1 to continue executing other transitions and it is P_2 to return to execute t_2 , and $P_1 + P_2 = 1$. The loop which contains t_2 has been executed s times, so when representing this subnet shown in Fig. 8a as b with subnet as p_c, t_c, p_{c+1} , the token in p_{c+1} arrives at:

$$\pi_{pc+1}(\tau) = (1 + \sum_{i=1}^s p_2^i) \times (\text{FTPC}(p_c) \oplus \text{FTTC}(t_1) \oplus d_{t1}(\tau)) \oplus \sum_{i=1}^s p_2^i \times (\text{FTPC}(p_2) \oplus \text{FTTC}(t_2) \oplus d_{t2}(\tau))$$

Fig. 8 Simplification of circle route



After Simplification, we can calculate the fuzzy time of each route in level k and then drag it to upper level $k - 1$ to calculate. After k times iteration until working out the fuzzy time level 0, it is temporal reachable probability of the whole workflow process. Through the stratification, calculation process is simplified, as well as the probability of state explosion is reduced.

4 Case Study and Analysis

A company receives an order and arranges for the procurement and production. Now, the concrete steps are shown as follows. (Fig. 9 shows its AFTWF-nets model of the business process, and Table 1 shows the significances and time information of transitions and places):

The order arrives at p_i at $\pi_{p_i}(\tau) = 1[0, 0, 0, 0]$, time calculation is shown as follows:

- Step 1: t_1 . As we know $e_{t_1}(\tau) = 1[0, 0, 0, 0]$, $FTTC(t_1) = 1[0, 0, 0, 0]$ so $o_{t_1}(\tau) = 1[0, 0, 0, 0]$.
- Step 2: p_1, p_2 and p_{12} . $p_1, p_2, p_{12} \in I_p(t_1)$ so $\pi_{p_1}(\tau) = \pi_{p_2}(\tau) = \pi_{p_{12}}(\tau) = o_{t_1}(\tau) \oplus d_{t_1}(\tau) = 1[2, 3, 5, 6]$.
- Step 3: t_2, p_3, t_4, p_5, t_6 and p_7 . $FTPC(p_1) = 1[0, 0, 0, 0]$, $FTTC(t_2) = 1[0, 0, 0, 0]$, $e_{t_2}(\tau) = 1[2, 3, 5, 6]$, so $o_{t_2}(\tau) = 1[2, 3, 5, 6]$, $\pi_{p_3}(\tau) = o_{t_2}(\tau) \oplus d_{t_2}(\tau) = 1[3, 5, 8, 10]$, $e_{t_4}(\tau) = \pi_{p_3}(\tau) \oplus FTPC(p_3) = 1[3, 5, 8, 10]$, $o_{t_4}(\tau) = e_{t_4}(\tau) \oplus FTTC(t_4) = 1[3, 5.5, 9, 11.5]$, and we get $\pi_{p_5}(\tau) = 1[4, 7.5, 12, 15.5]$, successively $e_{t_6}(\tau) = 1[4, 7.5, 12, 15.5]$, $o_{t_6}(\tau) = 1[4, 7.5, 12, 15.5]$, $\pi_{p_7}(\tau) = 1[12, 17, 23, 28]$.
- Step 4: p_2, t_3, p_4, t_5 and p_6 . $e_{t_3}(\tau) = \pi_{p_2}(\tau) \oplus FTPC(p_2) = 1[2, 3, 5, 6]$, so we get $o_{t_3}(\tau) = e_{t_3}(\tau) \oplus FTTC(t_3) = 1[2, 3, 5, 6]$ and $\pi_{p_4}(\tau) = o_{t_3}(\tau) \oplus d_{t_3}(\tau) =$

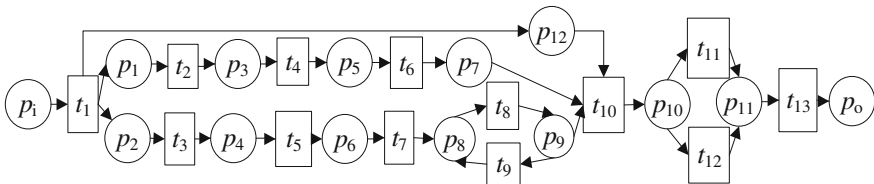


Fig. 9 AFTWF-nets model of business process

Table 1 Significances and time information of the transtions and places

t	Significance	$FTTC(t)$	$FD(t)$	p	Significance	$FTPC(p)$
t_1	Assignment	1[0, 0, 0, 0]	1[2, 3, 5, 6]	p_i	Process begins	1[0, 0, 0, 0]
t_2	Apply for funds	1[0, 0, 0, 0]	1[1, 2, 3, 4]	p_o	Process ends	1[0, 0, 0, 0]
t_3	Apply for funds	1[0, 0, 0, 0]	1[0, 1, 2, 3]	p_1	Local purchasing Dept	1[0, 0, 0, 0]
t_4	Appropriation	1[0, 0.5, 1, 1.5]	1[1, 2, 3, 4]	p_2	Subsidiary purchasing Dept	1[0, 0, 0, 0]
t_5	Appropriation	1[0, 0, 0, 0]	1[1, 2, 3, 4]	p_3	Local finance office	1[0, 0, 0, 0]
t_6	Purchase& deliver	1[0, 0, 0, 0]	1[8, 9.5, 11, 12.5]	p_4	Subsidiary finance office	1[0, 0, 0, 0]
t_7	Purchase& deliver	1[0, 0, 0, 0]	1[7, 8, 9, 10]	p_5	Funds	1[0, 0, 0, 0]
t_8	Quality check	1[0, 0, 0, 0]	1[1, 2, 3, 4]	p_6	Funds	1[0, 0, 0, 0]
t_9	Recheck	1[0, 0, 0, 0]	1[0, 0, 1, 1]	p_7	Raw material	1[0, 0, 0, 0]
t_{10}	Deliver production	1[0, 0, 0, 0]	1[0, 1, 1, 2]	p_8	Raw material	1[0, 0, 0, 0]
t_{11}	Produced by line A	1[1, 2, 2, 3]	1[8, 10, 12, 14]	p_9	Quality check Dept	1[0, 0, 0, 0]
t_{12}	Produced by line B	1[2, 3, 3, 4]	1[5, 6, 7, 8]	p_{10}	Production Dept	1[0, 0, 0, 0]
t_{13}	Deliver to users	1[0, 0, 0, 0]	1[0, 1, 2, 3]	p_{11}	Productions	1[0, 0, 0, 0]
				p_{12}	Warehouse	1[0, 5, 25, 30]

1[2, 4, 7, 9], and $e_{i5}(\tau) = 1[2, 4, 7, 9], o_{i5}(\tau) = 1[2, 4, 7, 9], \pi_{p6}(\tau) = 1[3, 6, 10, 13]$.

Step 5: t_7, p_8, t_8, p_9 and t_9 . Before loop $\pi_{p6}(\tau) = 1[3, 6, 10, 13], e_{i7}(\tau) = 1[3, 6, 10, 13]$ $o_{i7}(\tau) = 1[3, 6, 10, 13], e_{i8}(\tau) = 1[10, 14, 19, 23], o_{i8}(\tau) = 1[10, 14, 19, 23]$ and $\pi_{p9}(\tau) = 1[11, 16, 22, 27]$. Assume possibility that raw materials do not meet the standard, is 10 %, and when they are unqualified, t_9 is enabled, and then $e_{i9}(\tau) = 1[11, 16, 22, 27], o_{i9}(\tau) = 1[11, 16, 22, 27], d_{i9}(\tau) = 1[0, 0, 1, 1]$, so $\pi_{p8}(\tau)' = 1[11, 16, 23, 28], o_{i8}(\tau)' = 1[11, 16, 23, 28]$, and $\pi_{p9}(\tau)' = 1[12, 18, ; 26, 32]$. If circling once,

$$\pi_{p9}(\tau)'' = 90 \% \times \pi_{p9}(\tau) + 10 \% \times \pi_{p9}(\tau)'$$

$$(\tau)' = 1[11.1, 16.2, 22.4, 27.5]$$
.

Step 6: t_{10} $e_{i10}(\tau) = \text{latest}(\pi_{p7}(\tau), \pi_{p9}(\tau)'', \pi_{p12}(\tau)) = \min(1, 1, 1)[\max(12, 11.1, 0), \max(17, 16.2, 5), \max(23, 22.4, 25), \max(28, 27.4, 30)] = 1[12, 17, 25, 30]$, so $o_{i10}(\tau) = 1[12, 17, 25, 30], \pi_{p10}(\tau) = 1[12, 18, 26, 32]$.

Step 7: t_{11} and t_{12} . $e_{i11}(\tau) = e_{i12}(\tau) = 1[12, 18, 26, 32]$, so $o_{i11}(\tau) = \text{MIN}\{1[12, 18, 26, 32] \oplus 1[1, 2, 2, 3], \text{earliest}(1[12, 18, 26, 32] \oplus 1[1, 2, 2, 3], 1[12, 18, 26, 32] \oplus 1[2, 3, 3, 4])\} = \min\{1[13, 20, 28, 35], 1[13, 20, 28, 35]\} = 1[13, 20, 28, 35], o_{i12}(\tau) = \text{MIN}\{1[12, 18, 26, 32] \oplus 1[2, 3, 3, 4], \text{earliest}(1[12, 18, 26, 32] \oplus 1[1, 2, 2, 3], 1[12, 18, 26, 32] \oplus 1[2, 3, 3, 4])\} = \min\{1[14, 21, 29, 36], 1[13, 20, 28, 35]\} = 1[14,$

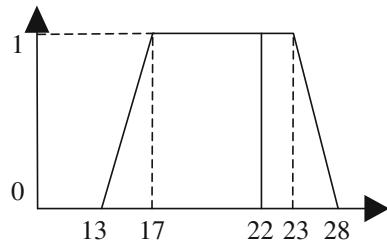
21, 28, 35], so if execute t_{11} , $\pi_{p11}(\tau) = o_{t11}(\tau) \oplus d_{t11}(\tau) = 1[21, 30, 40, 49]$, and if execute t_{12} , $\pi_{p11}(\tau)' = o_{t12}(\tau) \oplus d_{t12}(\tau) = 1[18, 26, 35, 43]$, and assume possibility that execute t_{11} is 60 %, that execute t_{12} is 40 %, so $\pi_{p11}(\tau)'' = 60 \% \times \pi_{p11}(\tau) + 40 \% \times \pi_{p11}(\tau)' = 1[19.8, 28.4, 38, 46.6]$.

Step 8: t_{13} and $p_{o.e113}(\tau) = 1[19.8, 28.4, 38, 46.6]$, $o_{t13}(\tau) = 1[19.8, 28.4, 38, 46.6]$ and $\pi_{po}(\tau) = 1[19.8, 29.4, 40, 49.6]$. It means that the total time cost is $1[19.8, 29.4, 40, 49.6]$.

After calculation, time performance of the workflow is shown as follows:

1. Activities execution time constraint satisfaction. For example, it is expected that local purchasing Dept should finish purchasing within 22 days, which means the token arrives at p_7 at $\pi_{p7}(\tau) = 1[22, 22, 22, 22]$. As shown in Fig. 10, the area of the left part of the trapezoidal is $[(22 - 17) + (22 - 13)] \times 1/2 = 7$, and the whole area is $[(23 - 17) + (28 - 13)] \times 1/2 = 10.5$, so the possibility that t_6 finishes within 22 days is $\text{Possibility}(\pi_{p7}(\tau) \leq \pi(\tau)) = 7/10.5 = 0.667$.

Fig. 10 $\pi_{p7}(\tau) \leq \pi(\tau)$



2. Time distance activities constraints satisfiability. There are some time constraints between production Dept receiving raw materials and delivering product to users. If time distance between t_{10} and t_{13} needs to be less than 5 days, it means to solve $\text{Possibility}(o_{t13}(\tau) \leq o_{t10}(\tau) \oplus 1[5, 5, 5, 5])$. As shown in Fig. 11,

Step 1:

$$\begin{aligned} &\text{Possibility}(o_{t13}(\tau) \leq o_{t10}(\tau) \oplus 1[5, 5, 5, 5]) \\ &= \text{Possibility}(1[19.8, 28.4, 38, 46.6] \leq 1[17, 22, 30, 35]) \end{aligned}$$

Fig. 11 $o_{t13}(\tau) \leq o_{t10}(\tau) \oplus 1[5, 5, 5, 5]$

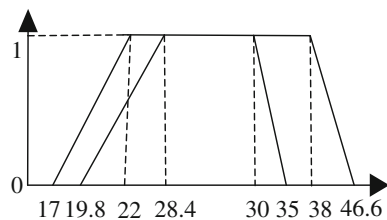
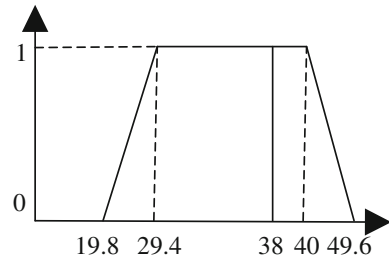


Fig. 12 $\pi_{po}(\tau) \leq \pi_f(\tau)$



- Step 2: The overlap area is $[(30 - 28.4) + (35 - 19.8)] \times 1/2 = 8.4$, the area of right half part is $[(38 - 28.4) + (46.6 - 19.8)] \times 1/2 = 18.2$, so Possibility($o_{t13}(\tau) \leq o_{t10}(\tau) \oplus 1[5, 5, 5, 5]$) = $8.4/18.2 = 0.462$,
- Step 3: That is the possibility that time distance is no more than 5 days.
3. Temporal reachable probability of the process. Firstly, calculate it using general method. From the above result, the total fuzzy time cost is $\pi_{po}(\tau) - \pi_{pi}(\tau) = 1[19.8, 29.4, 40, 49.6]$.

If the whole process finishes within 38 days, the expected fuzzy timestamp is $\pi_f(\tau) = 1[38, 38, 38, 38]$. As shown in Fig. 12, the area of the whole trapezoidal is $[(40 - 29.4) + (49.6 - 19.8)] \times 1/2 = 20.2$, the left part is $[(38 - 29.4) + (38 - 19.8)] \times 1/2 = 13.4$, so Possibility($\pi_{po}(\tau) \leq \pi_f(\tau)$) = $13.4/20.2 = 0.663$. Secondly, calculate it using stratification algorithm. HAFWF-nets is shown in Fig. 13 and the time information of each node is given above; therefore, we can obtain the time information of each basic route in level 3, substitute the result into the next upper level until finally substitute the result into level 0 and work out $\pi_{po}(\tau) = 1[19.8, 29.4, 40, 49.6]$, which is the same as the above answer.

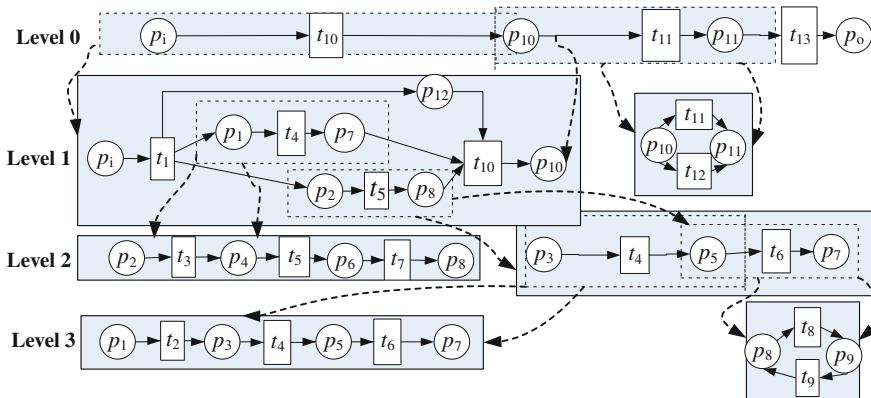


Fig. 13 HAFWF-nets model of business process

5 Conclusion

This paper proposed AFTWF-nets based on the existing research achievements on TCPN and FTHN and then gave its formal definition as well as corresponding reasoning and analysis algorithm. Firstly, AFTWF-nets has been used to describe the temporal information within workflows, and then model time constraints, and analyze its time performance. Secondly, based on the above analysis, a reduction algorithm for fuzzy time workflow model was put forward. Finally, through a practical interprovincial company application, the effectiveness of our method has been verified.

AFTWF-nets can be used to model those workflows which involved uncertain time information, and it will increase the flexibility of the workflow management systems, and also enrich workflow modeling theory and, promote the application of workflow management software.

References

1. Li H, Fan Y (2002) Overview on managing time in workflow systems. *J J Softw* 13(8):1552–1558
2. vander Aalst WMP (1998) The application of Petri nets to workflow management. *J J Circuits Syst Comput* 8(1):21–661
3. Ling S, Schmidt H (2000) Time Petri nets for workflow modeling and analysis. In: *Proceedings of IEEE international conference on systems, Man and Cybernetics*, IEEE Press, Nashville, pp 3039–3041
4. Li H, Fan Y (2004) Workflow model analysis based on time constraint Petri nets. *J J Softw*. 15(1):17–26
5. Murata T (1996) Temporal Uncertainty and fuzzy-timing high-level Petri nets. In: *Application and theory of Petri Nets (Lecture Notes in Computer Science)*, New York, pp 11–28
6. Pan Y, Tang Y (2006) Modeling of fuzzy temporal workflow nets and its time possibility analysis. *J Comput Integr Manuf Syst*. 12(11):1779–1784
7. Liu X, Yin G, Zhang Z (2009) A high-level fuzzy timed Petri-net-based approach. In: *2009 Fourth international conference on internet computing for science and engineering*, pp 131–136

Price Discount Subsidy Contract for a Two-Echelon Supply Chain with Random Yield

Weimin Ma, Xiaoxi Zhu and Miaomiao Wang

Abstract This paper revisits the traditional supplier–buyer integrated production–inventory model which deals with the problem of a manufacturer supplying a product to a retailer serving the end consumer with a known demand. We present two types of price discount subsidy, additional discount subsidy contract, and all units discount subsidy contract to coordinate the supply chain, and analytic solutions for each contract are described. We derived that with yield uncertainty, the manufacturer’s expected profit is concave on the manufacturer’s target production quantity under both contracts, and the discount rate is positively related to the manufacturer’s target production quantity under the all units discount subsidy contract. We derive the constraint condition under which the manufacturer will produce more under the discount subsidy contracts and which contracts are more available to promote the manufacturer’s expected output.

Keywords Random yield · Supply chain management · Price discount subsidy · Inventory control

1 Introduction

Uncertainty exists almost in every node in a supply chain. Stockout is one of uncertainties. The 2011 Japan earthquake disrupts the operation of the world semiconductor products supply chain. It might mean a loss of market share or a direct loss of profit for a retailer. Once a consumer goes to a store 2–3 times without finding the goods he wants, he may never go back to that store again. For a supplier, the lack of trust of the retailer may lead to a loss of the future demand and

W. Ma · X. Zhu (✉) · M. Wang
School of Economics and Management, Tongji University,
Shanghai, People’s Republic of China
e-mail: zhuxiaoxicome@163.com

latent order. Inaccuracy in the information system inventory as compared to the physical inventory may lead to out of stocks. Inaccuracy may occur for many reasons, a principal one brings randomness such as broken machines in a manufacturing system.

Traditional research [1–3] has shown that stockouts have a negative impact for retailers, both directly (on sales and profit) and indirectly (on customer satisfaction and retail image). Fitzsimons [1] suggests that stockout problems are not limited to traditional supermarkets, but may constitute a far more daunting problem for E-grocers who experience more severe forecasting problems and strongly fluctuating demand. Breugelmans et al. [4] investigates the impact of an online retailer's stockout policy on consumers' category purchase and choice decisions. Their results reveal that the adopted stockout policy has a significant impact on both decisions under an online grocery shopping experiment. Agrawal and Sharda [5] simulated a model which is employed to investigate the effect of such loss defined by the stock loss parameter (λ) and the frequent alignment of physical and information system inventories on the stockout (Sout) and average inventory (I). They indicated that there is a significant reduction in stockout loss when the alignment is done monthly versus annually, but it does not add much value beyond a monthly check. However, overproduction also does harm to the manufacturer for the occupation of funds and inventory. Offering discount is one of the best ways for the manufacturer to minimize his lost when overproduction occurs.

Many manufacturing plants need to produce before the orders that come from the downstream retailers because of that a factory's operation cannot stop when there is no order or there is no enough production capacity when big orders come. Such that there may exist an overproduction problem before the next order come. By considering the unexpected total holding costs, some manufacturers tend to sell the unsold products at a discounted price especially in some food industry like baker's shop, and almost all brands of garment enterprises sell at a discount price in the off-season for the costumes. Gurnani [6] study quantity discount pricing models with different ordering structures in a system consisting of a single supplier and heterogeneous buyers and show that order coordination always leads to a reduction in the system costs. Viswanathan and Wang [7] consider a single-vendor, single-retailer supply chain and evaluate the effectiveness of quantity discounts and volume discounts as coordination mechanisms in distribution channels with demand that is price sensitive. Pricing and quantity discount also provide coordination for supply chain management. Yue et al. [8] investigate the coordination of cooperative advertising in a two-level supply chain when manufacturer offers price deductions to customers. Qin et al. [9] consider volume discounts and franchise fees as coordination mechanisms in a system consisting of a supplier and a buyer with a price-sensitive demand. In their model, the supplier acts as the leader by announcing its pricing policy to the buyer in advance and the buyer acts as the follower by determining his unit selling price and annual sales volume.

Researchers have proposed several contracts that can coordinate the supply chain which has one supplier and multiple retailers; for example, pricing discount contracts (Sinha and Sarmah [10]; Xie et al. [11]), quantity discount contracts (Lau

et al. [12]; Krichen et al. [13]). However, they do not take the supply chain's production uncertainties into account.

However, oppositely the manufacturer's overproduction may be benefit for the retailer for the discounted selling products which are cheaper than the normal price. Especially, the kind of product which sell like hot cakes. Here, we consider a kind of product which its supply is always falls short of demand and the manufacturer usually cannot deliver the goods according to the ordered quantity on time for the yield randomness. Hence, the retailer is often in the situation of stockout, and the manufacturer will be faced with retailer penalty, and also, the manufacturer will suffer from holding costs when his final production quantity exceeds the retailer's demand. As the supplier and the leader of the supply chain, what is the production quantity the manufacturer should plan? What is the optimal order size the retailer should order to satisfy the end consumer's demand? In this paper, a two-echelon supply chain consists of a manufacturer and a retailer with a known and fixed demand and stochastic yield is explored. Yield risks are considered in the model. The manufacturer's optimal production quantity and the retailer's optimal order quantity are studied in this paper.

The single-period inventory problem has a variety of extensions, and other newsvendor-type problems have been introduced since the problem was first presented by Within [14]. Much effort has been spent on determining production and procurement when demand is uncertainty. Yet, less research has considered lot-sizing decisions with random yield. Yano and Lee [15] presented a thorough review of single-period models which dealt with lot-sizing models with random yields. Grosfeld and Ygerchak [16, 17] start with the multiple lot-sizing problem when the yield of each batch is random, and they provide a review of models, analytical results, and insights pertaining to multiple lot-sizing in production to order, with emphasis on multistage systems and inspection issues. He and Zhang [18] propose several risk sharing contracts that distribute the random yield risk among supply chain parties and evaluate the supply chain performances. But their work does not include the issue how to avoid the manufacturer's holding cost and guarantee the end consumer's demand under a random yield environment. In their study, a random defective rate is assumed and all defective items produced are reworked when regular production ends. Nowadays, due to the end customers' dynamic needs and demands, the retailer's order quantities and the manufacturer's production plan can change from time to time. Consequently, it is necessary to take consideration of the intention that making the manufacturer to produce no less than the retailer's order and finally meet the end consumer's demand in order to establish such contract that maximizes the manufacturer and meets the end consumer's demand. We propose two discount contracts to construct a bridge for coordinating the two objectives above. In practice, manufacturers choose his production quantity as the retailer ordered in order to avoid the inventory costs. But due to capacity or quality problems, the final output usually cannot meet the retailer's order. Thus, supply falls short of demand happens. The contract we proposed can be applied to settle this kind of problem by selling the manufacturer's unsold products at a discounted price to the retailer in order to release the

manufacturer's inventory cost and to satisfy the end consumer's demand with more assurance.

Following the introduction and literature review in Sect. 1, the rest of the paper is organized as follows. We present models, existence condition of the optimal production quantity of the manufacturer under random yield, and two hybrid contracts that urge the manufacture to promote his target production quantity, so that retailer's order can be met by a lager degree of assurance in Sect. 2. In Sect. 3, we formulate the retailer's decision problem by giving the optimal order quantity under two contracts under random yield and a numerical example is analyzed. Section 4 gives conclusions.

2 The Models

We consider a two-echelon supply chain (newsvendor) inventory problem in which the manufacturer sells products to the retailer. The product is a type a hot selling item that can be easily but not absolutely sold by the retailer. Here, we assume that there is no salvage cost for the unsold products. The manufacturer faces a fixed and known retailer demand D . The manufacturer charges the retailer a wholesale price ω per unit purchased. c is the product's unit production cost, and the production costs are taken into account even if the target unit is not converted to final yield due to randomness. h is the unit holding cost. π is per unit shortage cost (penalty cost) if the retailer's demand cannot be met, and it is non-negative. This part of revenue lost may include loss of reputation. We assumed that ω and π are predetermined by a contract when the actual demand is not placed by the retailer. Q stands for the manufacturer's target production quantity which is also the manufacturer's decision variable in our model. x presents the product's production risk which is a random variable. a, b , respectively, denote the lower bound and the upper bound of the random variable x . Define $f(x)$ as the probability density function of x , and in the examples, we assume $f(x)$ is uniformly distributed.

2.1 Additional Discount Subsidy Contract

Under this discount contract, the manufacturer may not suffer from the holding cost and salvage cost of over production, but abiding the lost of selling the products at a discount price. The holding cost saved for the manufacturer is a form of subsidy provided by the retailer, or the manufacturer will pay a lot for the holding cost. However, the retailer takes all risk of keeping any unsold stocks. The manufacturer do not need to worry about the holding cost of overproduction under this contract so that $h_1 = 0$ within this contract. This transaction between the manufacturer and the retailer is referred to as wholesale price and price discount contract. Here, we call this contract as discount subsidy contract for short. The real output is $x \cdot Q$, and the random variable x can be assumed that $0 \leq a \leq x \leq b$.

The manufacturer's net profit, which is also his objective function, is formulated in the following

$$\begin{aligned} \text{TP}_1(Q) = & \omega \cdot \min(xQ, D) + \beta_1 \omega \cdot \max[(xQ - D), 0] \\ & - \pi \cdot \max[(D - xQ), 0] - cQ. \end{aligned} \quad (1)$$

To obtain the expected total profit for the manufacturer in the following

$$\begin{aligned} E(\text{TP}_1(Q)) = & \omega Q \int_a^{D/Q} xf(x)dx - cQ + \beta_1 \int_{D/Q}^b (xQ - D)f(x)dx \\ & + \omega D \int_{D/Q}^b f(x)dx - \pi \int_a^{D/Q} (D - xQ)f(x) \\ = & Q(\omega + \pi) \int_a^{D/Q} xf(x)dx - cQ - \pi D \int_a^{D/Q} f(x)dx \\ & + \beta_1 Q \int_{D/Q}^b xf(x)dx + (\omega D - \beta_1 D) \int_{D/Q}^b f(x)dx \end{aligned} \quad (2)$$

To have the first-order condition of Q in (2), we get the optimal equation for the manufacturer

$$\frac{\partial E(\text{TP}(Q))}{\partial Q} = (\omega + \pi) \int_a^{D/Q^*} xf(x)dx + \beta_1 xf(x)dx - c. \quad (3)$$

To set $\partial E(\text{TP}_1(Q))/\partial Q = 0$, we have the optimal target quantity for production Q^* under additional discount subsidy contract is given

$$(\omega + \pi) \int_a^{D/Q^*} xf(x)dx + \beta_1 \int_{D/Q^*}^b xf(x)dx = c. \quad (4)$$

Proposition 1 *With yield randomness the manufacturer's expected profit is concave on Q^* under the additional discount subsidy contract.*

Proof 1 To have the second derivative of (3), we have

$$\frac{\partial^2 E(\text{TP}_1(Q))}{\partial Q^2} = -[\pi + (1 - \beta_1)\omega] \frac{D^2}{Q^3} f\left(\frac{D}{Q}\right) < 0. \quad (5)$$

Namely that there exist an optimal and unique Q^* which maximizes the manufacturer's net profit under the additional discount subsidy contract. To directly demonstrate the solution, we apply (2) for a special case where the production risk x is taken from a uniform distribution with probability density function $f(x) = 1/(b-a)$; $a \leq x \leq b$. The manufacturer's expected total profit for the case of uniform distribution production risk under additional discount subsidy contract is given by

$$E(\text{TP}_1(Q)) = (\omega + \pi) \frac{(D^2 - a^2Q^2)}{2(b-a)Q} - \frac{\pi D(D - aQ)}{(b-a)Q} + \beta_1 \frac{b^2Q^2 - D^2}{2(b-a)Q} + D(\omega - \beta_1) \frac{bQ - D}{(b-a)Q} - cQ. \tag{6}$$

The optimal target production quantity for the manufacturer is given by

$$Q^* = D \sqrt{\frac{\pi + (1 - \beta_1)\omega}{a^2(\omega + \pi) - \beta_1\omega b^2 + 2c(b - a)}}. \tag{7}$$

From (7), we can find that Q is proportional to the retailer's demand D (also his ordered quantity).

Proposition 2 *If $\mu > c/(\omega + \pi_1)$, the optimal manufacturer target production Q is a monotone increasing function of β_1 ; If $\mu < c/(\omega + \mu_1)$, the optimal manufacturer target production Q is a monotone decreasing function of β_1 .*

Proof 2 Equation (12) can be converted as

$$(\omega + \pi_1 - \beta_1\omega) \int_a^{D_1/Q} xf(x)dU = c - \beta_1\omega \int_a^b xf(x) \Rightarrow \int_a^{D_1/Q} xf(x) = \frac{c - \mu\beta_1\omega}{\omega + \pi_1 - \beta_1\omega}.$$

Let $\Delta = D_1/Q$, Δ is a constant which is decided by ω , π_1 , β_1 , and $xf(x)$. For $F(\xi) = \int_a^\xi Ug(U)dU$ is an increasing function of ξ . It is easy to obtain that Δ is monotone increasing function of c and decreasing function of penalty cost π_1 . Derive the first-order condition of Δ on β_1 , we have

$$\frac{\partial F(\Delta)}{\partial \beta_1} = \frac{\partial F(D_1/Q)}{\partial \beta_1} = \frac{-\mu\omega(\omega + \pi_1 + \omega c)}{[(1 - \beta_1)\omega + \pi_1]^2}$$

If $\mu > c/(\omega + \pi_1)$, Δ is a monotone increasing function of β_1 , and the optimal manufacturer target production Q is a monotone increasing function of β_1 . If $\mu < c/(\omega + \pi_1)$, namely that $\omega c - \mu\omega(\omega + \pi_1)$, Δ is a monotone decreasing function of β_1 . For Q decreased with Δ , thus, the optimal manufacturer target

production Q is a monotone decreasing function of β_1 . The manufacturer's best production quantity can be promoted by setting a larger β_1 in the contract.

Corollary 1 *Under uniform distributed yield risk, the additional discount contract would let the manufacturer to set a target production quantity larger than the retailer ordered when the wholesale price ω , penalty cost π , and discount rate β_1 satisfy*

$$(1 - a^2)(\omega + \pi) - \beta_1\omega(1 - b^2) > 2c(b - c) \quad (8)$$

Proof 3 It can be proved by solving the difference of Eq. (7) and retailer demand D

$$\begin{aligned} Q^* - D &> 0 \\ \Rightarrow D \left(\sqrt{\frac{\pi + (1 - \beta_1)\omega}{a^2(\omega + \pi) - \beta_1\omega b^2 + 2c(b - a)} - 1} \right) &> 0 \\ \Rightarrow \pi + (1 - \beta_1)\omega &> a^2(\omega + \pi) - \beta_1\omega b^2 + 2c(b - a) \\ \Rightarrow (1 - a^2)(\omega + \pi) - \beta_1\omega(1 - b^2) &> 2c(b - a). \end{aligned}$$

As the unit production cost c and the production risk x 's lower and upper bound a and b are fixed and already known to the supply chain member, the contract items ω , π , and β_1 should satisfy in Eq. (8) if the retailer wants the manufacturer to produce more.

2.2 All Units Discount Subsidy Contract

In order to prevent short delivery, the retailer may even want the manufacturer to produce more than he ordered. Here, we consider another incentive contract that no matter how many products the manufacturer produces the retailer will receive a fixed discount β_2 given in advance, including the condition of overproduction and stockout. The manufacturer's optimal production quantity is analyzed below. The net profit of the manufacturer is given by

$$\begin{aligned} \text{TP}_2(Q) = \beta_2\omega \cdot \{ \min(xQ, D) + \max[(xQ - D), 0] \} \\ - \pi \cdot \max[(D - xQ), 0] - cQ. \end{aligned} \quad (9)$$

The manufacturer's expected net profit is given by

$$\begin{aligned}
 E(\text{TP}_2(Q)) &= \beta_2\omega \left(Q \int_a^{D/Q} xf(x)dx + D \int_{D/Q}^b f(x)dx + \int_{D/Q}^b (xQ - D)f(x)dx \right) \\
 &\quad - \pi \int_a^{D/Q} (D - xQ)f(x)dx - cQ \\
 &= \beta_2pQ \int_a^b xf(x)dx - \pi \int_a^{D/Q} (D - xQ)f(x)dx - cQ.
 \end{aligned}
 \tag{10}$$

To derive the first-order condition of Q in (9), and we have

$$\frac{\partial E(\text{TP}_2(Q))}{\partial Q} = \beta_2\omega \int_a^b xf(x)dx + \pi \int_a^{D/Q} xf(x)dx - c
 \tag{11}$$

To set $\partial E(\text{TP}_2(Q))/\partial Q = 0$, we have the optimal target quantity for production Q_2^* under all units discount subsidy contract is given by

$$\beta_2\omega \int_a^b xf(x)dx + \pi \int_a^{D/Q^*} xf(x)dx = c.
 \tag{12}$$

To have the second derivative of (10), we have

$$\frac{\partial^2 E(\text{TP}_2(Q))}{\partial Q^2} = -\frac{D^2}{Q^3}f\left(\frac{D}{Q}\right) < 0.
 \tag{13}$$

Similar to lemma 1, we also find that there exist a Q^* which can maximize the manufacturer’s expected net profit. Namely that there exist an optimal and only one Q^* which will maximize the manufacturer’s net profit under the all units discount subsidy contract.

To directly demonstrate the solution, we apply (2) for a same case where the production risk x is taken from a uniform distribution with probability density function $f(x) = 1/(b-a)$; $a \leq x \leq b$. The expected total profit for the case of uniform distribution production risk under all units discount subsidy contract is given by

$$\text{ETP}_2(Q) = \beta_2\omega \frac{(a + b)Q}{2} - \frac{\pi(a^2Q^2 - D^2)}{2(b - a)Q} - \frac{\pi D(D - aQ)}{(b - a)Q} - cQ.
 \tag{14}$$

The optimal target production quantity for the manufacturer is given

$$Q^* = D \sqrt{\frac{\pi}{\pi a^2 - \omega \beta_2 (b^2 - a^2) + 2c(b - a)}} \tag{15}$$

From (7), we can find that Q is proportional to the retailer’s demand D (also his ordered quantity).

Proposition 3 *The discount rate β_2 is positively related to the manufacturer’s target production quantity Q^* under the all units discount subsidy contract. Namely that the manufacturer’s optimal target production quantity should be proportional to the discount rate β_2 .*

We also find that the product’s selling price P is proportional to the manufacturer target production quantity Q^* , and the units manufacture cost c has negative impacts on Q^* . By comparing the difference of Eqs. (7) and (15), we derive corollary 2.

Corollary 2 *Under uniform distributed yield risk, in order to have a larger manufacturer target production quantity, the retailer should select the all units discount contract when ω , π , β_1 , and β_2 satisfy*

$$\frac{\pi \beta_1 (b^2 - a^2) + 2c(\omega - \beta)(b^2 - a^2) + \omega \beta_2 (\beta_1 - \beta_2 - \beta_2 \pi)(b^2 - a^2)}{[\pi a^2 - \omega \beta_2 (b^2 - a^2) + 2c(b^2 - a^2)] \cdot [a^2(\omega + \pi) - \beta_1 b^2 + 2c(b^2 - a^2)]} > 0. \tag{16}$$

Proof 4 It can be proved by solving the difference of Eqs. (15) and (7) as done.

3 The Retailer’s Problem

In our research, the end consumer does not directly buy products from the manufacturer. The direct customer of the manufacturer is the retailer. The retailer buys the products and sells them to the end consumer. The retailer decides his order size by considering the end consumer’s demand, the stochastic production capacity, the holding costs, and the shortage costs, and at the same time, the manufacturer will make his production decision to optimize his objective profit function. In order to distinguish between the manufacturer and the retailer, we mark the manufacturer’s parameters with the index 1 and the retailer’s parameters with the index 2.

We formulate the retailer’s problem of finding the optimal order quantity as a mathematical programming problem where the manufacturer’s problem of finding the optimal target production quantity is a constraint. On modeling the retailer’s situation, we add several parameters: P stands for the end consumer’s unit buying price of the product from the retailer. h_2 is the unit holding cost for the retailer. π_2

is the unit penalty cost of the retailer for not meeting the end consumer’s demand. The following sections are focus on the retailer’s decision problem.

3.1 The Retailer–Manufacturer Problem Under Additional Discount Subsidy Contract

The objective profit function of the retailer $RTP_1(Q)$ under the additional discount subsidy contract is

$$\begin{aligned}
 RTP_1(Q) = & p \cdot \min(xQ, D_2) - \omega \cdot \min(xQ, D_1) - \beta_1 \omega \cdot \max[(xQ - D_1), 0] \\
 & - h \cdot \max[(xQ - D_2), 0] - \pi_2 \cdot \max[(D_2 - xQ), 0] \\
 & + \pi_1 \cdot \max[(D_1 - xQ), 0]
 \end{aligned} \tag{17}$$

The manufacturer’s mathematical programming problem is to maximize the expected net profit, taking into consideration the manufacturer’s optimal target production quantity as a constraint. The retailer’s decision model is formulated as

$$\begin{aligned}
 E[RTP_1(Q)] = & pQ \int_a^{D_2/Q} xf(x)dx + pD \int_{D_2/Q}^b f(x)dx - \omega Q \int_a^{D_1/Q} xf(x)dx \\
 & - \omega D_1 \int_{D_1/Q}^b f(x)dx - \beta_1 \int_{D_1/Q}^b (xQ - D_1)f(x)dx \\
 & - \pi_2 \int_a^{D_2/Q} (D_2 - xQ)f(x)dx - h_2 \int_{D_2/Q}^b (xQ - D_2)f(x)dx \\
 & + \pi_1 \int_a^{D_1/Q} (D_1 - xQ)f(x)dx \\
 \text{st. } & (\omega + \pi_1) \int_a^{D_1/Q} xf(x)dx - \beta_1 \int_{D_1/Q}^b xf(x)dx = c
 \end{aligned} \tag{18}$$

We also assume here that the production risk x is taken from a uniform distribution. With this assumption, we substitute a uniform probability density function in (16) and turn it into

$$\begin{aligned}
 \text{Max}_{D_1} E[\text{RTP}_1(Q)] &= \frac{(p + \pi_2)Q}{2(b - a)} \left(\frac{D_2^2}{Q^2} - a^2 \right) + \frac{(p + h_2)D_2}{(b - a)} \left(b - \frac{D_2}{Q} \right) \\
 &\quad - \frac{(\omega + \pi_1)Q}{2(b - a)} \left(\frac{D_1^2}{Q^2} - a^2 \right) - \frac{(\omega - \beta_1)D_1}{(b - a)} \left(b - \frac{D_1}{Q} \right) \\
 &\quad - \frac{\beta_1 Q}{2(b - a)} \left(b^2 - \frac{D_1^2}{Q^2} \right) - \frac{\pi_2 D_2}{(b - a)} \left(\frac{D_2}{Q} - b \right) \\
 &\quad - \frac{h_2 Q}{2(b - a)} \left(b^2 - \frac{D_2^2}{Q^2} \right) + \frac{\pi_1 D_1}{(b - a)} \left(\frac{D_1}{Q} - a \right) \\
 \text{st. } Q &= D_1 \sqrt{\frac{\omega + \pi_1 - \beta_1}{a^2(\omega + \pi_1) - \beta_1 b^2 + 2c(b - a)}}
 \end{aligned} \tag{19}$$

Substitute Q into the manufacturer’s objective function and we get

$$D_1^* = D_2 \sqrt{\frac{p + \pi_2 + h_2}{A^2 [b^2(h_2 + \beta_1) - a^2(\omega + \pi_1 - p - \pi_2)] + 2A(a\pi_1 - b\beta_1 + b\omega) - (\omega + \beta_1 + \pi_1)}} \tag{20}$$

where $A = \sqrt{\frac{\omega + \pi_1 - \beta_1}{a^2(\omega + \pi_1) - \beta_1 b^2 + 2c(b - a)}}$

D_1^* is the optimal lot size for the retailer to maximize his expected net profit.

3.2 The Retailer–Manufacturer Problem Under all Units Discount Subsidy Contract

The objective profit function of the retailer $\text{RTP}_2(Q)$ under the additional discount subsidy contract is

$$\begin{aligned}
 \text{RTP}_2(Q) &= p \cdot \min(xQ, D_2) - \beta_2 \omega \cdot \{ \min(xQ, D) + \max[(xQ - D), 0] \} \\
 &\quad - h \cdot \max[(xQ - D_2), 0] - \pi_2 \cdot \max[(D_2 - xQ), 0] \\
 &\quad + \pi_1 \cdot \max[(D_1 - xQ), 0]
 \end{aligned} \tag{21}$$

The manufacturer’s mathematical programming problem is to maximize the expected net profit, taking into consideration the manufacturer’s optimal target production quantity as a constraint. The retailer’s decision model is formulated as

$$\begin{aligned}
 \text{Max}_{D_1} E[\text{RTP}_2(Q)] = & pQ \int_a^{D_2/Q} xf(x)dx + pD \int_{D_2/Q}^b f(x)dx \\
 & - \beta_2\omega \left(Q \int_a^{D_1/Q} xf(x)dx + D \int_{D_1/Q}^b f(x)dx \right. \\
 & \quad \left. + \int_{D_1/Q}^b (xQ - D_1)f(x)dx \right) \\
 & - \pi_2 \int_a^{D_2/Q} (D_2 - xQ)f(x)dx \\
 & - h_2 \int_{D_2/Q}^b (xQ - D_2)f(x)dx \\
 & + \pi_1 \int_a^{D_1/Q} (D_1 - xQ)f(x)dx \\
 \text{st. } & \beta_2\omega \int_a^b xf(x)dx + \pi_1 \int_a^{D_1/Q} xf(x)dx = c
 \end{aligned} \tag{22}$$

We also assume here that the production risk x is taken from a uniform distribution. With this assumption, we substitute a uniform probability density function in (20) and turn it into

$$\begin{aligned}
 \text{Max}_{D_1} E[\text{RTP}_2(Q)] = & \frac{p_2Q}{2(b-a)} \left(\frac{D_2^2}{Q^2} - a^2 \right) + \frac{p_2D_2}{(b-a)} \left(b - \frac{D_2}{Q} \right) \\
 & - \beta_2p_1 \left(\frac{Q}{2(b-a)} \left(\frac{D_1^2}{Q^2} - a^2 \right) + \frac{D_1}{(b-a)} \left(b - \frac{D_1}{Q} \right) \right) \\
 & \quad \left(+ \frac{Q}{2(b-a)} \left(b^2 - \frac{D_1^2}{Q^2} \right) - \frac{D_1}{(b-a)} \left(b - \frac{D_1}{Q} \right) \right) \\
 & + \frac{\pi_2Q}{2(b-a)} \left(\frac{D_2^2}{Q^2} - a^2 \right) - \frac{\pi_2D_2}{(b-a)} \left(\frac{D_2}{Q} - a \right) \\
 & - \frac{h_2Q}{2(b-a)} \left(b^2 - \frac{D_2^2}{Q^2} \right) + \frac{h_2D_2}{(b-a)} \left(b - \frac{D_2}{Q} \right) \\
 & - \frac{\pi_1D_1}{(b-a)} \left(\frac{D_1}{Q} - a \right) + \frac{\pi_1Q}{2(b-a)} \left(\frac{D_1^2}{Q^2} - a^2 \right) \\
 \text{st. } & Q = D_1 \sqrt{\frac{\pi_1}{\pi_1a^2 - \beta_2p_1(b^2 - a^2) + 2c(b-a)}}
 \end{aligned} \tag{23}$$

Substitute Q into the manufacturer’s objective function and we get

$$D_1^* = D_2 \sqrt{\frac{B(p + \pi_2 + h_2)}{\pi_1 - B^2[\pi_2 a^2 - pa^2 - \omega\beta_2(b^2 - a^2) - hb^2 - \pi_1 a^2]}} \tag{24}$$

where $B = \sqrt{\frac{\pi_1}{\pi_1 a^2 - \beta_2 \omega (b^2 - a^2) + 2c(b-a)}}$

D_1^* is the optimal lot size for the retailer to maximize his expected net profit.

Numerical example

In order to directly express our model, we further assume that $\omega = 1.32$, $c = 0.62$, $\pi_1 = 0.33$, $\beta_1 = 0.8$, $a = 0.9$, $b = 1.03$, $D_1 = 318,000$. Table 1 shows the comparison of the two contracts.

Under additional discount subsidy contract, without any optimization, the retailer would set to $D_1 = D_2 = 318,000$. This trivial solution forces the manufacturer to plan $Q_1^* = 363,930$ with an expected profit of $E(TP_1) = 219,860$, and with an expected retailer profit of $E(RTP_1) = 1,515,000$. However, with optimization, according to our model and as shown in Table 1, the retailer sets his order quantity $D_1 = 347,540$ (29,540 units over the demand), and the manufacturer’s expected net profit is 240,280 (+9.29 % more than the trivial solution); 50,200 (397,740–347,540) units product are sold to the retailer at a discounted price rate 0.8. Here, we defined an “average” selling price as $P_a = [(Q^* - D_1) \cdot \omega\beta_1 + \omega D_1] / Q^*$ and the “average” selling price under the additional discount subsidy contract $P_{a1} = 1.2867$. With the optimization of D_1 , the retailer obtains an expected profit $E(RTP_1) = 1,376,800$ (–9.12 % more than the trivial solution). We may conclude that the optimization of D_1 under additional discount subsidy contract may increase the expected net profit for the manufacturer but decrease the expected revenue for the retailer. The reason will be illustrated in the next paragraph. The end user also may gain some benefit since larger D_1 increases the probability for supplying all his needs.

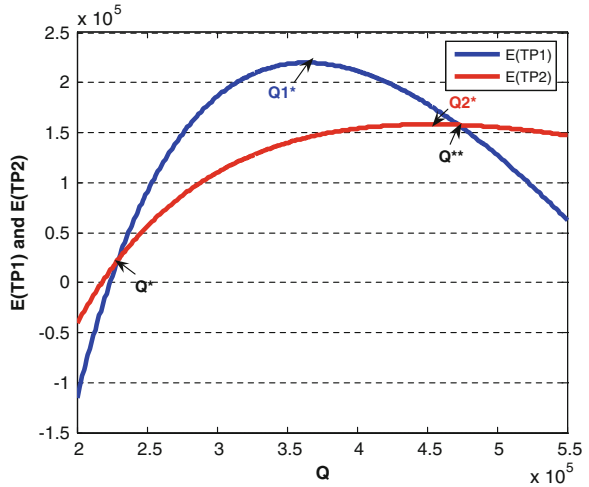
Under all units discount subsidy contract, the trivial solution forces the manufacturer to plan $Q_1^* = 451,710$ with an expected profit of $E(TP_1(Q)) = 158,220$ and with an expected retailer profit of $E(RTP_2(Q)) = 845,700$. But with optimization, the retailer sets $D_1 = 297,740$ (20,260 units below the demand), and the manufacturer plans $Q_1 = 422,930$ with expected profit $E(TP_1) = 148,140$ (–6.37 % more than the trivial solution); 148,140 units product are sold to the retailer at a discounted rate 0.8. The “average” selling price under the all units discount subsidy contract $P_{a2} = \beta_2 \cdot \omega = 1.0560$. With the optimization of D_1 , the retailer obtains an expected profit $E(RTP_2(Q)) = 1,089,100$ (+28.78 %). We may conclude that the optimization of D_1 under all units discount subsidy contract may decrease the expected profit of the manufacturer but increase the retailer’s expected profit. The end consumer may also gain satisfaction since larger D_1 increases the probability for supplying all his demand.

The reason why the manufacturer’s expected profit is increased under the additional subsidy contract but decreased under the all units discount subsidy contract with the optimization of D_1 can be illustrated as the lower “average”

Table 1 The comparison of the two contracts

Classification	Additional Discount Subsidy Contract			All Discount Subsidy Contract			
	Q_{o}^*	$E(\pi_{o}^M(Q_o^*))$	$D_{2,o}^*$	Q_d^*	$E(\pi_d^M(Q_d^*))$	$D_{2,d}^*$	$E(\pi_d^R(Q))$
Retailer with optimization	363930	219860		422930	158220	318000	1089100
Retailer with no optimization	397740	240280	347540	422930	148140	297740	1089100

Fig. 1 Expected profits as a function of Q^* and Q^{**}



selling price ($1.0560 < 1.2867$) under the additional discount subsidy contract comparing to all units discount subsidy contract.

So, which contract is better? We use Fig. 1 to analyze which contract is better for the manufacturer.

The corresponding target production quantity of the two intersections is $Q^* = 229,287.7$ and $Q^{**} = 472,443.3$. At intersection Q^* and Q^{**} , the manufacturer has same expected net profit 22,501 and 157,650 under two discount subsidy contracts. The manufacturer prefers to choose the additional discount subsidy contract when his target quantity is within the interval (Q^*, Q^{**}) , because of that within the interval, the manufacturer could have a higher expected revenue comparing to choose all units discount subsidy contract. Thus, it is optimal to choose all units discount subsidy contract when his target quantity is out of the interval (Q^*, Q^{**}) .

4 Conclusions

Production planning and inventory control activities are often undertaken in uncertain environments. The well-known case as Dell’s zero-inventory policy operates successfully. However, it is difficult for manufacturing companies with random yield to have a zero inventory. An improper level of inventory may directly influence an enterprise’s survival and development. In this paper, we analyze and solve a special case of a single-period inventory problem with manufacturer random yield by proposing two contracts. Two types of discount subsidy contract are presented for the manufacturer–retailer supply chain in which the manufacturer does not have to worry about the holding cost and the retailer’s order is satisfied with more assurance. We derive that with yield uncertainty, the

manufacturer's expected profit $E(TP(Q))$ is concave on his target production quantity under both contracts, and the discount rate b_2 is positively related to Q^* under the all units discount subsidy contract. We derive the constraint condition under which Q^* can be promoted by the discount subsidy contracts, and the condition of which contracts is more available to promote Q^* .

For both types of contract, we showed that the retailer's problem can be formulated as a mathematical programming problem, where the optimum solution of the manufacturer is a constraint. The models that were developed here can be applied at both the manufacturer level and the retailer level. Maximizing the expected profits and satisfying the consumer demand are the two objectives for a supply chain. We derive that it is optimal for both of the manufacturer and the retailer to choose the additional discount subsidy contract when they prefer a larger expected profits. Yet if meeting the end consumer's demand is more important, they should choose the all units discount contract. In reality, which contract would be signed is determined by the "stronger" side in the supply chain.

In this paper, we propose two contracts to coordinate the supply chain. However, there may exist some other contracts like buyback and revenue sharing contract in a supply chain, and it is necessary to take these contracts into account to extend our model. Demand uncertainty is another important factor for a manufacturing system, and the problem becomes more complex when the consumer's demand is unknown to the retailer.

Acknowledgments The work was partly supported by the National Natural Science Foundation of China (71071113), a Ph. D. Programs Foundation of Ministry of Education of China (20100072110011), and the Fundamental Research Funds for the Central Universities.

References

1. Fitzsimons GJ (2000) Consumer response to stockouts. *J Consum Res* 27:249–266
2. Campo K, Gijsbrechts E, Nisol P (2003) The impact of retailer stockouts on whether, how much and what to buy. *Int J Res Mark* 20:273–286
3. Sloot LM, Verhoef PC, Franses PH (2005) The impact of brand equity and the hedonic level of a product on consumer stockout reactions. *J Retail* 81:15–34
4. Breugelmans E, Campo K, Gijsbrechts E (2006) Opportunities for active stockout management in online stores: The impact of the stockout policy on online stockout reactions. *J Retail* 82:215–228
5. Agrawal PM, Sharda R (2012) Impact of frequency of alignment of physical and information system inventories on out of stocks: a simulation study. *Int J Prod Econ* 136:45–55
6. Gurnani H (2001) A study of quantity discount pricing models with different ordering structures: Order coordination, order consolidation, and multi-tier ordering hierarchy. *Int J Prod Econ* 72:203–225
7. Viswanathan S, Wang Q (2003) Discount pricing decisions in distribution channels with price-sensitive demand. *Eur J Oper Res* 149:571–587
8. Yue JF, Austin J, Wang MC et al (2006) Coordination of cooperative advertising in a two-level supply chain when manufacturer offers discount. *Eur J Oper Res* 168:65–85

9. Qin YY, Tang HW, Guo CH (2007) Channel coordination and volume discounts with price-sensitive demand. *Int J Prod Econ* 105:43–53
10. Sinha S, Sarmah SP (2010) Single-vendor multi-buyer discount pricing model under stochastic demand environment. *Comput Ind Eng* 59:945–953
11. Xie JX, Zhou DM, Wei JC et al (2010) Price discount based on early order commitment in a single manufacturer-multiple retailer supply chain. *Eur J Oper Res* 200:368–376
12. Lau AHL, Lau HS, Zhou YW (2008) Quantity discount and handling-charge reduction schemes for a manufacturer supplying numerous heterogeneous retailers. *Int J Prod Econ* 113:425–445
13. Krichen S, Laabidi A, Abdelaziz FB (2011) Single supplier multiple cooperative retailers inventory model with quantity discount and permissible delay in payments. *Comput Ind Eng* 60:164–172
14. Within TM (1995) Inventory control and price theory. *Manage Sci* 2:61–80
15. Yano C, Lee HL (1995) Lot sizing with random yields: a review. *Oper Res* 43:311–334
16. Gerchak Y, Grosfeld-Nir A (1998) Multiple lot-sizing, and value of probabilistic information, in production to order of an uncertain size. *Int J Prod Econ* 56–57:191–197
17. Grosfeld-Nir A, Gerchak Y (2004) Multiple lotsizing in production to order with random yields: review of recent advances. *Ann Oper Res* 126:43–69
18. He Y, Zhang J (2008) Random yield risk sharing in a two-level supply chain. *Int J Prod Econ* 112:769–781

Reducing EMI in a PC Power Supply with Chaos Control

Yuhong Song, Zhong Li, Junying Niu, Guidong Zhang,
Wolfgang Halang and Holger Hirsch

Abstract A chaos control methodology has been proposed for suppressing EMI in DC–DC converters in the former research. In this paper, this method is applied in the power supply ATX as a real practical application, where an external chaotic signal is generated in a control circuit for pulse width modulation (PWM). As compared with the conventional PWM, this methodology has prominent effectiveness in suppressing EMI by spreading the power spectra smoothly over the frequency band.

Keywords Switched-mode power supply · Chaos control · EMI

1 Introduction

A switched-mode (or switching-mode) power supply (SMPS) incorporates a switching regulator to convert electrical power with high efficiency. The switching power makes use of high rates of changes in voltage and current, resulting in the electromagnetic interference (EMI), which impairs other devices' performance and harms human being's health.

EMI is illustrated by the large peaks in the power spectrum, which locate at the operation frequency and its harmonics, and can be tackled through the spread

Y. Song (✉) · J. Niu
Department of Electronic and Information Engineering, Shunde Polytechnic, Shunde, China
e-mail: syhscut@163.com

Y. Song · Z. Li · J. Niu · G. Zhang · W. Halang
Faculty of Mathematics and Computer Science, FernUniversität, Hagen, Germany

H. Hirsch
Faculty of Engineering, University of Duisburg-Essen, Duisburg-Essen, Germany

spectrum by means of chaos control, due to the pseudo-random and continuous spectrum characteristics of chaos.

Hong Li et al. [1, 2] have contributed to the application of chaos control in DC–DC converters to reduce EMI, including the system design, dynamic analysis, simulations, and hardware implementations of chaotic DC–DC converters. There, the models, such as boost converters, are ideal, lacking a practical application. Therefore, a real example, ATX 2.0 power supply, is to be given to verify the effectiveness of chaos control in EMI suppression.

2 Design of a Power Supply with Chaotic PWM

Select a practical product, ATX 2.0 of PC power supply, which is a typical accessory of desktop computer for a nominal input voltage 220 V AC and a typical valid value from 200 to 240 V/3.5 A by 50 Hz. The supply produces 200 W power and has multiple DC outputs, that is, +3.3 V, +5 V, –5 V, +12 V, –12 V, and +5 V standby voltage output (SB).

2.1 System Circuit

ATX 2.0 power supply system is composed of many modules, which are depicted in Fig. 1. AC Input flows through the EMI filters and then into input the rectifier and filter, out of which DC 310 V normally comes, which operates on the half-bridge reverting module. From the PWM control circuit, two pulse signals come out for driving two high-frequency (HF) transistors, which turn on and off in turn. At the second part of HF isolating transformer, there are multiple outputs crossing the rectifier and filter. Finally, from the circuit come out multiple DC voltages. There are some other components such as the auxiliary source, signal detection, and protection parts. The main circuit consists of a half-bridge reverting module

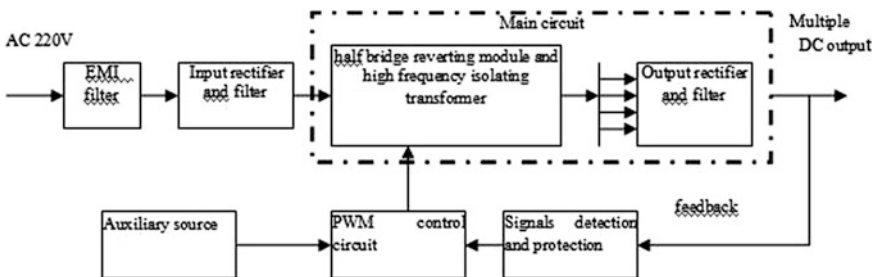


Fig. 1 Circuit diagram

and an HF isolating transformer, which can generate multiple DC voltages, such as $\pm 5\text{ V}$, $\pm 12\text{ V}$. The main circuit is shown in the dashed rectangle in Fig. 1.

2.2 Analog Chaotic PWM with TL494

The above-mentioned PWM control module is applied in a special IC TL494 [3], which is a fixed-frequency pulse width modulation control circuit, incorporating the primary building blocks required for the control of a switching power supply, as shown in Fig. 2. An internal linear sawtooth oscillator is frequency programmable by two external components, R_T and C_T . The oscillator frequency is determined by

$$f_{osc} = \frac{1.1}{R_T \cdot C_T} \tag{1}$$

The control signals are external inputs that can be fed back into the dead time control (DTC) through Pin 4, and the inputs of the error amplifier come through Pins 1, 2, 15, and 16, or the feedback input through Pin 3, as shown in Fig. 2. Output pulse width modulation is accomplished by comparing the positive sawtooth waveform across the capacitor, C_T , with either of non-inverting inputs of the dead time comparator and the PWM comparator. When the sawtooth wave transient value is greater than that of the control signals, the NOR gates are enabled only if the flip-flop clock pulse, denoted by CK, changes to logic 0. NOR gates are used to drive output transistors Q1 and Q2. Therefore, an increment in the control signal amplitude causes a corresponding decrease in the output pulse width.

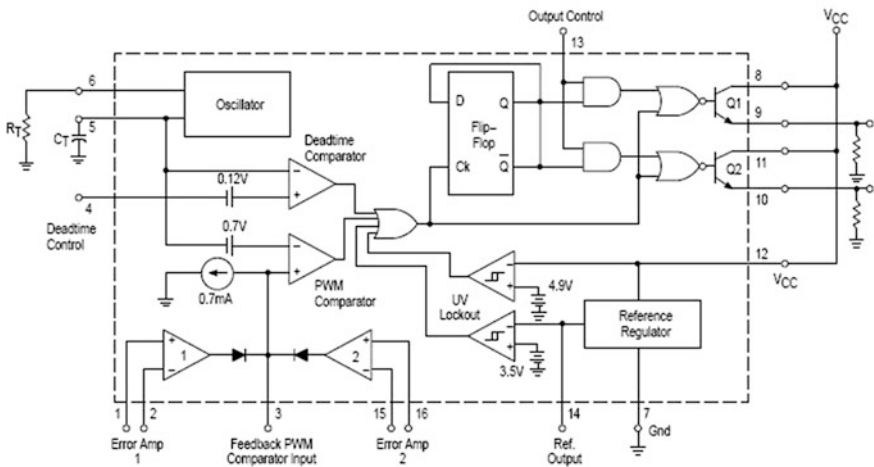


Fig. 2 TL494 block diagram

Assume that the output control at Pin 13 keeps at logic 1. Denote the outputs of error amplifiers 1 and 2 by u_{e1out} and u_{e2out} , respectively, the signal at Pin 5 by u_{ct} , another input of the DTC comparator by u_{dead} , and the feedback signal at Pin 3 by u_{fdb} .

Fix the DTC input u_{dead} to a constant between 0 and 3.3 V and assume that u_{e1out} and u_{e2out} are constant, so the outputs of TL494 are determined by u_{dead} , u_{ct} and u_{fdb} .

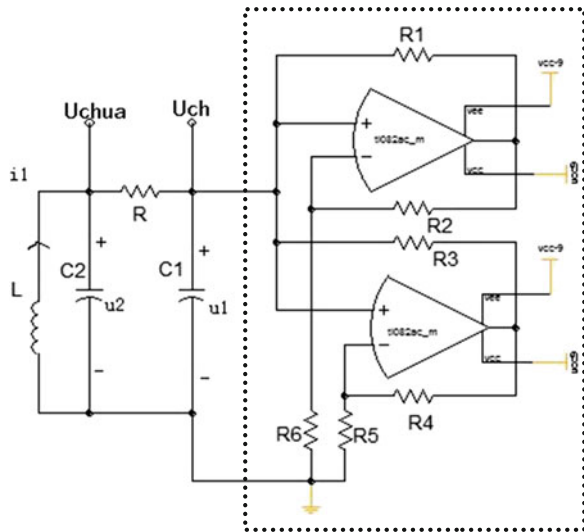
2.3 Design of an External Chaotic Signal to TL494

In recent decades, chaotic oscillators have been extensively investigated [4–6]. Chaotic oscillators have been used in the PWM control of DC–DC converters to reduce EMI [7]. Among the existing chaotic oscillators, Chua’s, Lorenz’s, and Chen’s oscillators are well known. In this paper, Chua’s oscillator is adopted due to its simplicity and maturity. As shown in Fig. 3, the Chua’s diode is depicted in the dotted line block Chua’s oscillator can be described as

$$\begin{cases} \frac{du_1}{dt} = \frac{1}{C_1} [(u_2 - u_1)G - f(u_1)] \\ \frac{du_2}{dt} = \frac{1}{C_2} [(u_1 - u_2)G + i_1] \\ \frac{di_1}{dt} = -\frac{1}{L}(u_2 + r_0 i_1) \end{cases} \quad (2)$$

where G stands for the reciprocal of ohm (Ω) to Chua’s diode and r_0 the resistor of the inductor L .

Fig. 3 The Chua’s circuit



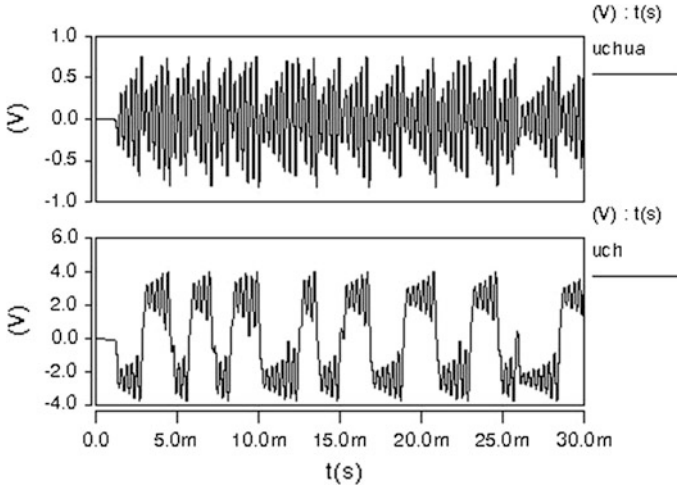


Fig. 4 Transient voltage waveforms

For instance, when $R = 1780 \Omega$, $R_1 = R_2 = 220 \Omega$, $R_3 = R_4 = 22 \text{ K } \Omega$, $R_5 = 3.3 \text{ K } \Omega$, $R_6 = 2.2 \text{ K } \Omega$, $L = 18 \text{ m henry (H)}$, $C1 = 100 \text{ n farad (F)}$, $C2 = 10 \text{ n F}$, the signals of the chaotic oscillator are shown in Fig. 4.

The voltage at the feedback Pin 3 of TL494, u_{fdb} , is required to vary between 0.5 and 3.5 V; thus, an amplifier is designed to scale the signal $u_{chua} = u_2$ into [0.5, 3.5]. The circuit output is indicated by $u_{chaotic}$. The amplifier can be described by the following equations:

$$\begin{cases} u_+ = u_- \\ \frac{u_{chua} - u_+}{R_{11}} + \frac{12 - u_+}{R_{14}} = \frac{u_+}{R_{12}} \\ \frac{u_-}{R_{13}} = \frac{u_{chaotic} - u_-}{R_f} \end{cases} \quad (3)$$

Furthermore, the output signal of the amplifier can be described by the following equation:

$$u_{chaotic} = \frac{(R_f + R_{13}) * (12R_{11}R_{12} + R_{12}R_{14}u_{chua})}{R_{13} * (R_{11}R_{14} + R_{11}R_{12} + R_{12}R_{14})} \quad (4)$$

where $R_{11} = R_{12} = R_{13} = 10 \text{ K } \Omega$, $R_{14} = 30 \text{ K } \Omega$, $R_f = 7 \text{ K } \Omega$, as shown in Fig. 5. The waveform diagram is shown in Fig. 6. It is remarked from (4) that $u_{chaotic}$ is a linear transition of u_{chua} .

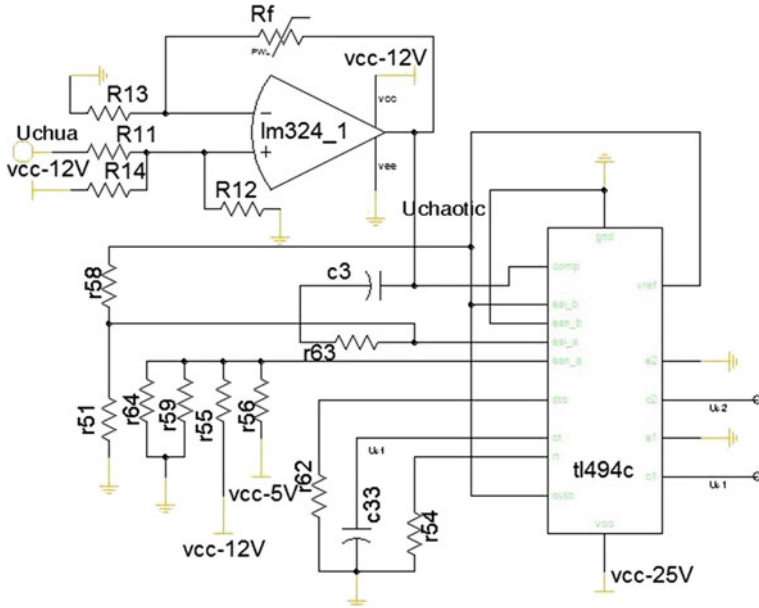
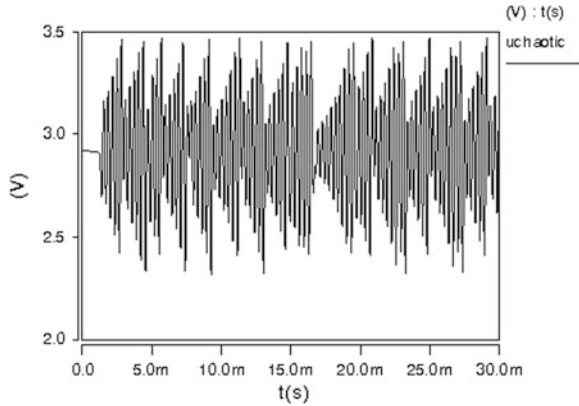


Fig. 5 Chaotic PWM module

Fig. 6 Amplifier output waveform



2.4 Chaos Control Based on TL494

Chua’s circuit (shown in Fig. 3) generates an chaotic voltage, u_{chua} , across the capacitor C_2 , which is analog. The output, $u_{chaotic}$, which is processed by the amplifier, is proportional and translational to u_{chua} (in Fig. 5). So $u_{chaotic}$ holds the frequency properties of u_{chua} . Their fundamental frequency is about 3 K hertz (Hz).

The oscillator frequency of TL494 is determined by (1). Once R_T and C_T (shown in Fig. 2) are fixed, the PWM has a constant frequency. The sawtooth waveform u_{ct} denotes the voltage across C_T , where $R_T = 12 \text{ K } \Omega$ and $C_T = 1.5 \text{ n F}$ f_{osc} can be obtained as 61 K Hz. Imposing a chaotic signal to Pin 3 results in

$$u_{fdb} = u_{chaotic} \quad (5)$$

When $u_{ct} > u_{dead}$, and $u_{ct} > u_{chaotic} - 0.7$, one has

$$CK = 0, Q1_b = \bar{Q}n, \text{ and } Q2_b = Qn$$

otherwise,

$$CK = 1, Q1_b = 0, Q2_b = 0$$

where Qn and $\bar{Q}n$ stand for the outputs of the flip-flop and have opposite states. If Qn is at logic 1, then $\bar{Q}n$ is at logic 0.

It is remarked that a change in chaotic signal amplitude causes a corresponding change in the output pulse width. That is, chaotic pulse width modulation (CPWM) holds the frequency, while the pulse width of the CPWM varies chaotically.

3 Simulation Results

A default value to the DTC is given by connecting the resistor to the ground, and inputting fixed values to two error amplifiers results in two zero outputs. Let $R_T = 12 \text{ K } \Omega$ and $C_T = 1.5 \text{ n F}$, and the parameters of other components are taken from a real ATX 2.0 power supply.

By simulating with Saber version 2008, a comparison is made between operating in conventional PWM model and operating in chaotic PWM model of using external chaos generator. EMI reduction is observed by using FFT by spreading the spectra of the critical component signals such as the voltage between the collector and emitter of the HF transistor, the voltage across the HF transformer, and the inductor currents of the multiple outputs. The simulation results are given in Fig. 7.

It is seen from Fig. 7 that, comparing to the conventional PWM control, the amplitude at the multiples of the baseband is greatly cut down in chaos control, implying the reduction in EMI.

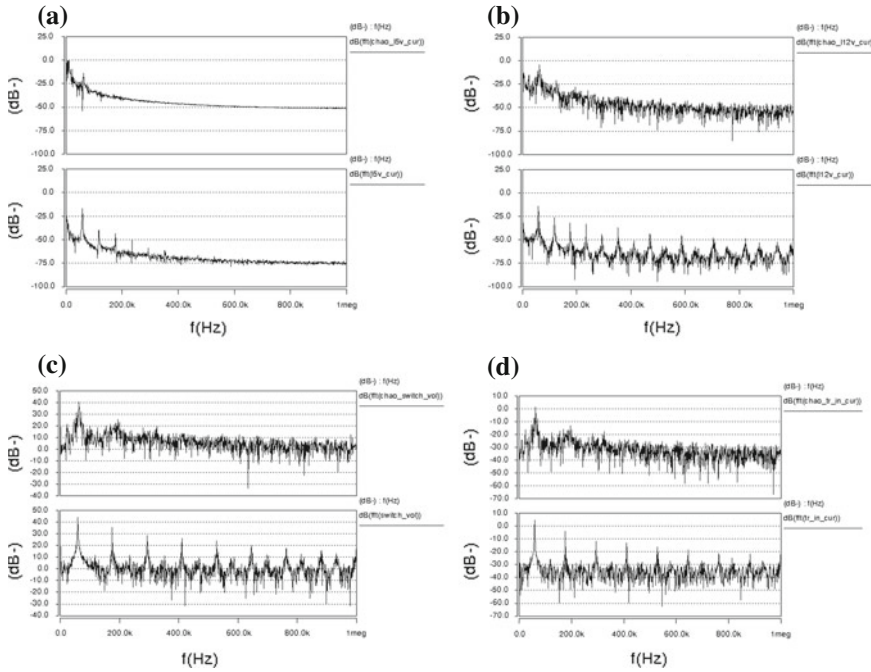


Fig. 7 FFT spectra(lower waveform of every picture corresponds to the conventional PWM control and upper to the chaotic PWM control). **a** Inductor current at +5 V output. **b** Inductor current at +12 V output. **c** Use of HF transistor. **d** Primary winding current of HF transformer

4 Hardware Design and Experimental Results

To further verify the effectiveness of the analog chaotic PWM, experimental and hardware designs have also been modeled. The key is to design the external module, which is composed of two parts: one is the circuit generating chaotic signals and another is to process the chaotic signals to suit for the IC TL494.

To generate chaotic signals, an improved Chua’s circuit is designed. So far, many methods have been reported to build Chua’s diode [8], among which the most popular one is shown in the dotted line block in Fig. 8. The parameter design is given in [9]. Here, the parameters for Chua’s diode are chosen as $R_{d1} = 2.4 \text{ K } \Omega$, $R_{d2} = 3.3 \text{ K } \Omega$, $R_{d3} = R_{d4} = 220 \text{ } \Omega$, and $R_{d5} = R_{d6} = 20 \text{ K } \Omega$. In order to realize oscillation in the experiment, the LC oscillator has not been adopted because the parameters of the inductor L are difficult to tune. As shown in Fig. 4.1, two amplifiers tl082c have been adopted. Other parameters of the Chua’s oscillator are set as $R_1 = R_2 = R_3 = R_4 = 2 \text{ K } \Omega$, $C_1 = 10 \text{ n F}$, $C_2 = 100 \text{ n F}$, $C_3 = 4.7 \text{ n F}$, and $R = 1.78 \text{ K } \Omega$.

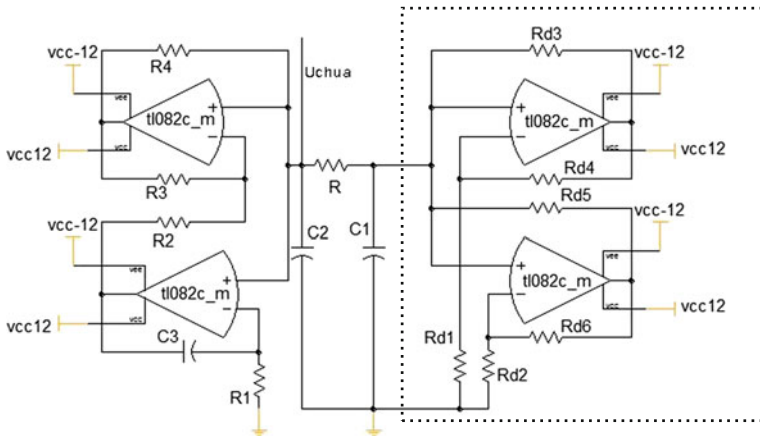


Fig. 8 Improved Chua’s circuit

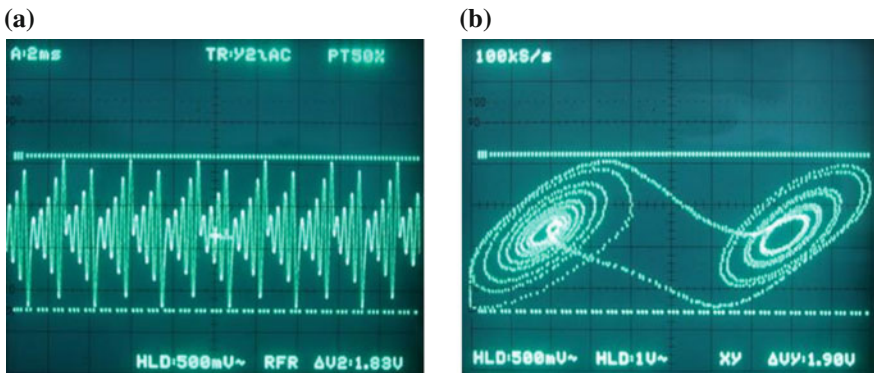


Fig. 9 Waveform of improved Chua’s circuit. **a** Transient voltage across capacitor C2. **b** The phase portrait of $V_{c2}-V_{c1}$

The experimental waveforms are shown in Fig. 9.

To meet the TL494 circuit, chaotic signal u_{chua} has to be processed by an amplifier. Then, u_{chua} becomes positive by a translation with a certain value, but it is still chaotic, that is, $u_{chaotic}$ keeps the frequency properties of u_{chua} . The circuit is shown in Fig. 10. It is known that $u_{chaotic}$ is the linear transition of u_{chua} . For $R_{11} = R_{12} = R_{13} = 10\text{ K } \Omega$, $R_{14} = 30\text{ K } \Omega$, and $R_f = 7\text{ K } \Omega$, the experimental waveforms are shown in Fig. 11.

In this experiment, the EMC standards en55022VM_B (QP class B) and en55022VM_B (AV class B) are applied, the measurement bandwidth is 9 kHz, the frequency step 5 kHz, the attenuation 10 dB, and the frequency range 0.15–30 MHz. The equipment under test is neither housing nor EMI filters.

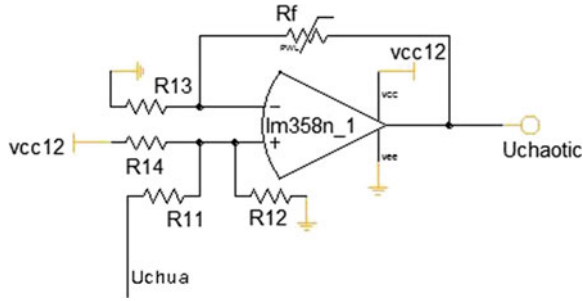


Fig. 10 The amplifying circuit

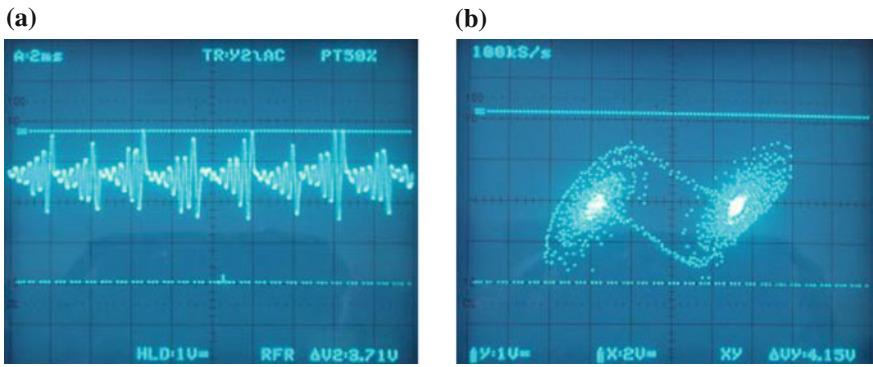


Fig. 11 Waveforms of the amplifying circuit. a Transient voltage of the output. b The phase portrait of $V_{chaotic}-V_{c1}$

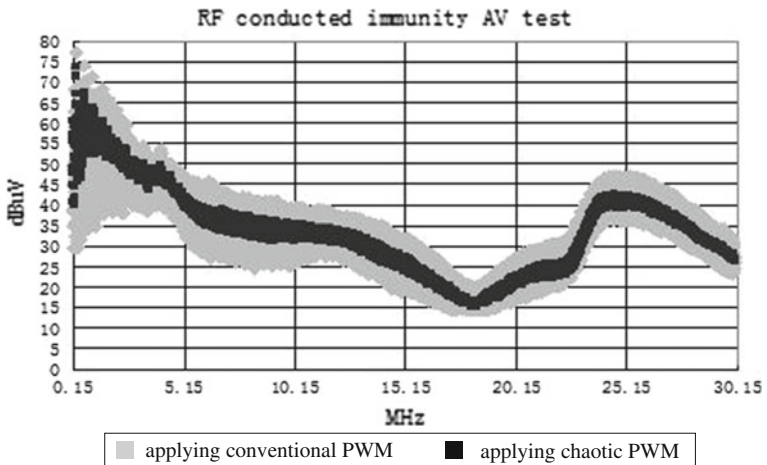


Fig. 12 Comparison of EMI

As shown in Fig. 12, gray curve is AV test result with conventional PWM control and black curve is the respective result with chaotic PWM control. In whole test frequency scope, the level of the latter is under the envelope curve of the former reduced by about 5 dBuV.

5 Conclusion

Chaos control has been proposed to suppress EMI in a practical power supply. Employing an external chaotic signal to a control circuit for PWM in ATX 2.0 power supply for PC implies a real application of chaos control for suppressing EMI. Simulation and experimental results have shown the effectiveness of the proposed methodologies for EMI suppression.

Acknowledgment This work was supported by AiF project under grant no. 17211 N in German.

References

1. Li H, Li Z, Zhang B, Tang WKS, HaLang W (2009) Suppressing electromagnetic interference in direct current converters. *IEEE Circuits Syst Mag* 2:10–28, Fourth quarter
2. Li H (2009) Reducing electromagnetic interference in DC–DC converters with chaos control. *Fortschritt -Berichte VDI*
3. Switchmode Pulse Width Modulation Control Circuit (2012) <http://www.datasheetcatalog.org/datasheet/motorola/TL494.pdf>
4. Bilotta E, Pantano P, Stranges F (2007) A gallery of chua attractors: Part I. *Int J Bifurcat Chaos* 17:1–60
5. Chen G (1993) Controlling Chua's global unfolding circuit family. *IEEE Trans Circuits Syst Part I* 40:829–832
6. Chua LO, Komuro M, Matsumoto T (1986) The double scroll family. part i: rigorous proof of chaos. *IEEE Trans Circuits Syst* 33:1072–1096
7. Balestra M, Lazzarini M, Setti G, Rovatti R (2005) Experimental performance evaluation of a low -EMI chaos-based current-programmed dc–dc boost converter. In: *Proceedings of IEEE international symposium circuits and systems*, vol 2. pp 1489–1492, May 2005
8. Chua LO (2007) <http://www.scholarpedia.org/article/Chua> circuit
9. Elwakil AS, Kennedy MP (2000) Chua's circuit decomposition: a systematic design approach for chaotic oscillators. *J Franklin Inst* 337:251–265

Optimization of Plasma Fabric Surface Treatment by Modeling with Neural Networks

Radhia Abd Jelil, Xianyi Zeng, Ludovic Koehl and Anne Perwuelz

Abstract Artificial neural networks (ANNs) are used to model the relationship between plasma processing parameters and woven fabric surface wetting properties. In this model, fourteen features characterizing woven structures and two plasma parameters are taken as input variables and the water contact angle cosine and capillarity height as output variables. In order to reduce the complexity of the model, a fuzzy logic-based method is used to select the most relevant parameters that are taken as inputs of the reduced neural model. Two techniques (early stopping and Bayesian regularization) are used for improving the generalization ability of neural networks. A methodology for optimizing such models is described. Moreover, a connection weight method is used to investigate the relative importance of each input variable. From the experiments, we find a good agreement between experimental and predicted data and that Bayesian regularization technique is the most suitable to achieve a good generalization.

Keywords Artificial neural networks · Fuzzy logic-based selection criterion · Industrial modeling · Atmospheric air plasma · Woven fabrics · Wettability

1 Introduction

In recent years, atmospheric pressure plasma has gained a growing interest for applications in the textile industry, due to its low cost, fast line speeds, and high reactivity of plasma gas-phase species. The energetic species in gas plasma like ions,

R. Abd Jelil (✉)

Institut Supérieur des Arts et Métiers de Tataouine, 3200 Tataouine, Tunisia
e-mail: abdjelilradhia@yahoo.fr

X. Zeng · L. Koehl · A. Perwuelz

The ENSAIT Textile Institute, 59056 Roubaix, France
e-mail: xianyi.zeng@ensait.fr

electrons, radicals, metastable, and UV photons can perform numerous surface modification processes such as surface activation, cleaning, grafting, and etching without altering material bulk properties [1]. Some works [2–4] have studied the effect of single factors on surface wetting properties. However, combinational effects of multiple factors have not been studied systematically. These factors such as electrical power, treatment speed, fabric nature, air permeability, surface roughness, and peak density have different effects on fabric surface wetting properties. The relationship between these factors and surface wetting properties is very complex and nonlinear. Thus, we use neural networks to construct a model. In fact, neural networks have numerous attractive properties for modeling complex systems such as efficient learning from experimental data and universal approximation for any arbitrary relations, and capacity of adaptation to imperfect or incomplete data [5].

In general, the modeling of industrial process is often constrained by complex process structure, large dimensionality of the input space, presence of redundant variables, and lack of available learning data. These factors may cause a deterioration of the generalization ability and an increase in the computational cost. Therefore, selecting the most relevant input variables is critical to build an appropriate and reduced model with high performance and improve the interpretability of the results and related model structure. In the literature, many feature selection methods have been proposed and studied [6–9]. In our study, as the number of available experimental data is rather limited, we use the fuzzy logic-based sensitivity variation criterion developed by Deng et al. [9] to select the most relevant fabric parameters to plasma treatment effects on fabric surfaces. By comparison with other numerical criteria, the proposed method has shown to be more robust, more efficient for physical interpretation, and less sensitive to measured data noises and uncertainties. Furthermore, it can deal with a small number of learning data. These advantages prove a strong motivation to the present paper for using such method to select the most relevant plasma process parameters in order to reduce model complexity.

In this paper, feed-forward artificial neural networks (ANNs) techniques are used for modeling the relationship between selected relevant plasma process parameters and fabric surface wetting properties, including the water contact angle cosine and the capillarity height of woven fabrics. The early stopping and Bayesian regularization techniques are used for improving the generalization ability. A connection weight approach is used to evaluate the relative importance of each input variable in the model. The performance of the model is analyzed and validated with experimental data.

2 Plasma Treatment

Plasma treatments are carried out using an atmospheric plasma machine called “Coating star” manufactured by the Ahlbrandt System Company (Fig. 1.). The following machine parameters are kept constant: a frequency of 30 kHz, an

electrode length of 0.5 m, and an inter-electrode distance of 1.5 mm. The varying process factors include the electrical power and treatment speed. Plasma discharge is generated at atmospheric pressure by two electrodes and a counter-electrode, both covered by a dielectric ceramic material. During plasma treatment, woven samples are in contact with the counter-electrode and passed through the plasma gas present between the electrodes and counter-electrode gap.

In order to quantify the surface treatment modification, wettability and capillarity measurements are carried out using distilled water (as liquid) on a “3S balance” from GBX Instruments. During measurements, a vertically hanging woven fabric sample of size 5 × 3 cm is connected to the “3S balance” at the weighing position and progressively brought into contact with the surface of water placed in a container. On immediate contact with the water surface, a sudden increase in weight is measured due to meniscus formation on the fabric surface. The weight increases further as the liquid flows inside the fabric structure by capillarity. Let the duration of each measurement to be 2 min. As soon as the liquid is detached from the fabric sample by moving the liquid container downwards, the balance gives the values of the total weight at the end (W_t) and the weight of capillarity (W_c). These two parameters are used to calculate the approximate meniscus weight (W_m) according to Eq. (1).

$$W_m = W_t - W_c \tag{1}$$

The cosine of the water contact angle ($\cos \theta$) of woven samples could be determined from the calculated meniscus weight using Eq. (2), since both the surface tension of liquid water and the perimeter of the contacting surface were known [4].

$$W_m g = p \gamma_L \cos \theta \tag{2}$$

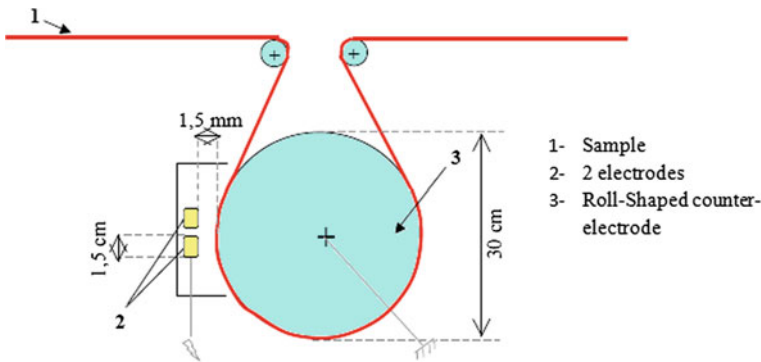


Fig. 1 Plasma treatment under atmospheric pressure by means of dielectric barrier discharge, using “Coating Star” plasma machine

where p is the sample perimeter in contact with the liquid (mm), W_m the calculated meniscus weight (g), $g = 9.81 \text{ m/s}^2$, γ_L the surface tension of the liquid (mN/m), and θ the contact angle ($^\circ$).

The capillarity height values for the woven samples were deduced from the capillarity weight values (W_c) using Eq. (3)

$$W_c = \rho_l h S \quad (3)$$

where h is the capillarity height that corresponds to the vertical distance reached by the liquid after 2 min of contact (mm), S the surface of the pores perpendicular to the flow of the liquid (mm^2), and ρ_l the liquid volume weight (g/cm^3).

3 Selection of Relevant Input Variables

In this paper, the fuzzy logic-based sensitivity variation criterion developed by Deng et al. [9] is used to select the most relevant input parameters of plasma process. The main advantage of this method is that it can deal with a small number of learning data. Its principle consists of calculating distances or variations between individual data samples in the input space (process parameters) and the output space (quality features), respectively. Then, fuzzy logic is used to evaluate the sensitivity variation of each input variable related to the output variable. The sensitivity for all the input variables is defined according to the two following principles:

If a small variation of an input variable Δx corresponds to a large variation of the output variable Δy , then this input variable has a great sensitivity value S .

If a large variation of an input variable Δx corresponds to a small variation of the output variable Δy , then this input variable has a small sensitivity value S .

These principles are transformed into a fuzzy model in which the input data variation Δx and the output data variation Δy , are taken as two input variables, respectively, and the sensitivity S as output variable [10].

Given a specific output variable y_l , for any pair of data sample (x_i, y_{jl}) and (x_j, y_{jl}) denoted as (i, j) , the input data variation Δx_{ij} and the output data variation Δy_{ij} are calculated. The corresponding sensitivity in the data pair (i, j) related to y_l can be obtained from this fuzzy model, that is, $S_l(i, j) = FL(\Delta x_{ij}, \Delta y_{ij})$.

When removing x_k from the whole set of input variables, the sensitivity of the remaining input variables in the data pair (i, j) related to the output y_l can be calculated by $S_{k,l}(i, j) = FL(\Delta x_{ij}^k, \Delta y_{ij})$. The sensitivity variation of the pair (i, j) can be calculated as follows:

$$\Delta S_{k,l}(i, j) = \left| FL(\Delta x_{ij}, \Delta y_{ij}) - FL(\Delta x_{ij}^k, \Delta y_{ij}) \right| \quad (4)$$

The general sensitivity variation $\Delta S_{k,l}$ for all pairs of data samples when removing the variable x_k is defined by:

$$\Delta S_{k,l} = \frac{1}{\gamma} \sum_{i=1}^n \sum_{j=i+1}^n \Delta S_{k,l}(i, j) \tag{5}$$

where $\gamma = n(n - 1)/2$ the total number of data pairs.

Bigger is the value of $\Delta S_{k,l}$, more the corresponding variable x_k is relevant to the quality feature y_l .

Based on this fuzzy logic-based sensitivity variation criterion, we proposed the following algorithm for selecting the most relevant variables.

Inputs: process input variables $X = \{x_1, \dots, x_k, \dots, x_m\}$, and one related specific output y_l

Output: relevant process parameters X_r and related values of sensitivity variation ΔS

$\text{corr}(x_i, x_k)$ denotes correlation between x_i and x_k

Initialize $X' = X, X_r = \{\}, \Delta S'_l = \{\}$

While $X' \neq \Phi$

Calculate the sensitivity variation of inputs in X' related to y_l denoted

$$\Delta S'_l = \{\Delta S_{1,l}, \dots, \Delta S_{k,l}, \dots, \Delta S_{\text{size}(X),l}\}$$

$$X_r = X_r \cup \{x_i\}, X' = X' \setminus \{x_i\} \text{ where } \Delta S_{i,l} = \max(\Delta S'_l)$$

$$X' = X' \setminus \{x_j, x_k\} \text{ where } \Delta S_{j,l} = \min(\Delta S'_l) \text{ and } \text{corr}(x_i, x_k) \geq 0.8$$

End

$$\Delta S = \Delta S'$$

This algorithm combines both the forward and the backward search. At each step, it removes the most sensitive variable, the most insensitive variable, and subsequently the variables that are in correlation with the most sensitive one. The variable which sensitivity variation is maximal is considered as the most sensitive variable. The most insensitive variable corresponds to the case in which its sensitivity variation is minimal. When this recurrent procedure is completed, we could obtain a significant and independent list of the most relevant process parameters.

4 Modeling with a Neural Network

4.1 Data Pre-processing

Neural network modeling will be made more efficient if a suitable data pre-processing procedure is performed. Before neural network training, it is often useful to scale the input and output data so that they all fall within a specified range. In our case, all the input and output data are scaled so as to have a normal distribution with zero mean and unity standard deviation using Eq. (6)

$$\text{Scaled value} = \frac{(\text{Actual value} - \mu)}{\sigma} \quad (6)$$

where μ and σ are the mean and standard deviation of the actual data, respectively.

The entire fabric samples with different production parameters are randomly divided into two subsets: 102 samples (75 %) for training and 34 samples (25 %) for test. Whenever early stopping technique is used, the initial training set is divided in the same way into a training set (68 samples) and a validation set (34 samples). After finishing the training, the outputs of the network must to be post-processed in a similar way to convert them to physically meaningful quantities.

4.2 Neural Network Architecture

In this study, a feed-forward neural network was used for the plasma modeling due to its proven high accuracy in learning nonlinear process data [5, 11]. The network architecture consisted of three layers of neurons: input layer, hidden layer, and output layer. The input layer corresponded to the selected input parameters. The outputs were the water contact angle cosine and the capillarity height values. In this application, two different activation functions were used: a sigmoid transfer function in the hidden layer and a linear transfer function in the output layer (Fig. 2).

4.3 Training Algorithms

In our research, we study two training algorithms for improving the network generalization ability: the Levenberg–Marquardt algorithm (*trainlm*) and the Bayesian regularization algorithm (*trainbr*). The first algorithm uses an early stopping mechanism in which the error on the validation and test set is monitored during the training process and training is stopped when the prediction accuracy begins to decrease, whereas the second algorithm is a modification of the first one to improve the model's generalization capability. The modification consists of

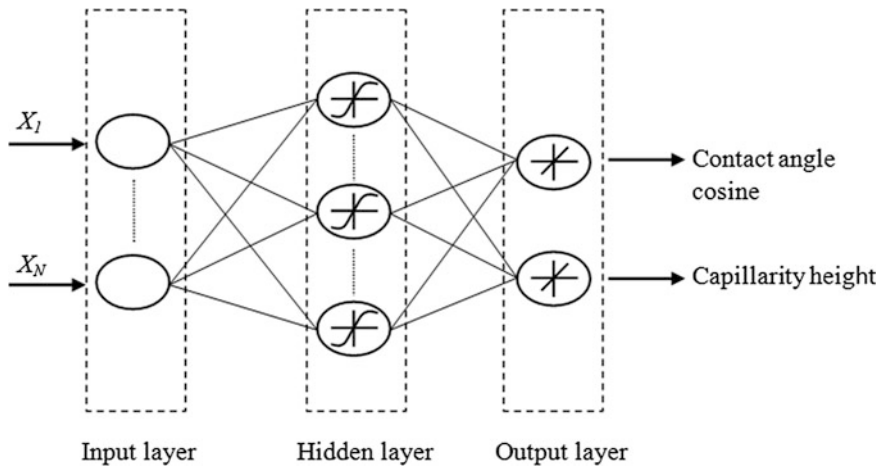


Fig. 2 Structure of the neural network used in this study. $X_1 \dots X_N =$ input vector

changing the performance function, which is normally chosen to be the sum of squares of the network errors (MSE), by adding a term that consists of mean square error of weights and biases. This performance function will cause the network to have smaller weight and biases, thereby forcing networks less likely to be overfit.

4.4 Optimizing the Number of Hidden Neurons

The number of hidden neurons is one of the key elements for designing a neural network [12]. If it is too small, the network cannot learn the problem correctly. If it is too large, that will lead to overfitting and may increase training time. When overfitting occurs, the network will began to model random noise in the data. Thus, it is extremely important to select an appropriate number of neurons in the hidden layer in order that the effective error in both training and testing drops to an acceptable value. Up to now, no systematic rules or equations lead to the optimal number of hidden neurons [13]. In many applications, this number is selected by trial and error. In this study, we propose an algorithm to determine the optimal number of neurons in the hidden layer. The principle of this algorithm is to first generate a network having one neuron in the hidden layer and then add neurons one by one recurrently until some stopping criteria are reached. The samples are split into three sets, namely training (50 %), validation (25 %), and testing (25 %). The validation set is used to terminate training done with the training set. The testing set is kept independent and used in accuracy assessment only after training has converged. This algorithm can be illustrated using the following pseudo-codes,

1. Create an initial network with one neuron in the hidden layer $N_H = 1$.
2. Realize 50 training iterations.
3. Calculate the root mean square errors ($RMSE_{Training}$, $RMSE_{Test}$) and the correlation coefficients ($R_{Training}$, R_{Test}) in the training and test sets.
4. Compare $RMSE_{Test}$ and $RMSE_{Training}$ and then compare the correlation coefficients to 1.

IF $\left(\frac{RMSE_{Test}}{RMSE_{Training}} < \alpha\right)$ and $(R_{Training} > 0.85)$ and $(R_{Test} > 0.85)$, THEN update the model with N_H neurons
 ELSE
 Let $N_H = N_H + 1$ and go back to Step 2
 END IF

According to this algorithm, 50 iterations are applied each time and two criteria are defined to determine when the insertion of new hidden neurons can be stopped. As neural network is an alternate statistical method, the root mean square error (RMSE) and the correlation coefficient (R) can be used as performance criteria leading to more suitable model structure. In fact, RMSE is an indicator of overall approximation accuracy, whereas R value is a measure of the strength of the relationship between predicted and real measurements. Here, the number of hidden neurons is considered optimal when the training and test errors are both in the same order and as small as possible, and the correlation coefficients are close to 1. The test and training RMSEs are considered in the same order if their ratio is close to 1. Therefore, this ratio should be less than a given threshold value α to obtain good network's generalization ability. Overall, this method will help to find the optimal or at least the near-optimal number of hidden neurons since the learning algorithms used can avoid being trapped into local minima. In our applications, α is set to be 1.5. The training and test RMSEs are calculated according to Eqs. (7) and (8), respectively,

$$RMSE_{Training} = \sqrt{\frac{1}{N_T} \sum_{i=1}^{N_T} (d_i - y_i)^2} \quad (7)$$

$$RMSE_{Test} = \sqrt{\frac{1}{N_t} \sum_{i=1}^{N_t} (d_i - y_i)^2} \quad (8)$$

where N_T is the number of training samples, N_t the number of test samples, d_i the desired output, and y_i the calculated output of the network.

The values of R are obtained by calculating the regression coefficients of the lines that relate network output values to their corresponding targets. In this application, R values superior to 0.85 are considered as good matching to the targets.

4.5 Number of Training Iterations

The number of iterations (i.e., epochs) is very important to assess the ability of neural network model. In general, this number is set by the user. In this subsection, we propose an algorithm to determine the optimal number of training iterations that allows improvement efficiency of the network model obtained in the previous subsection. The samples are split in the same manner as in the previous subsection. This algorithm includes the following pseudo-codes.

1. Set the numbers of hidden neurons to the optimal number (N_H) obtained by using the previous algorithm.
2. Begin with n_0 training iterations $n_{itr} = n_0$.
3. Calculate the RMSEs and the correlation coefficients in the training and test sets.
4. Compare the RMSEs and the correlation coefficients to those of the predictive model obtained previously.

```

IF the RMSEs and the  $R$  values of the network model are improved
THEN update the model with  $n_{itr}$  iterations
ELSE
Let  $n_{itr} = n_{itr} + k$  and go back to Step 3
END IF

```

In this algorithm, the number of training iterations is incremented from n_0 by k for each time. This incremental procedure stops when both the model's errors and R values are improved. It should be noted that the numbers n_0 and k can be chosen randomly to be adapted to the process complexity. In our applications, n_0 and k are set to be 50.

5 Results and Discussion

In this study, fourteen woven fabric features and two plasma parameters are taken as input parameters of the plasma process. These parameters are pre-selected by experts according to their possible influence on the outputs, that is, the cosine of the water contact angle and the capillarity height, as shown in Table 1. If we present all these 16 input parameters to the neural network, this would increase the network size, which leads to an increase in the amount of data required to estimate connection weights efficiently and decreases the processing speed. In order to reduce the size of the network and improve model performance, we use the fuzzy logic-based method presented forward in this paper to select the relevant input variables and remove irrelevant ones. Table 2 shows the detailed steps for recursively selecting the inputs relevant to the water contact angle cosine.

Table 1 Inputs and outputs of the plasma process

Inputs and outputs	Variables and their significance
Plasma process parameters (inputs)	Woven fabric features Composition (x_1), weave construction (x_2), fiber count (x_3), fabric weight (x_4), thickness (x_5), weft density (x_6), warp density (x_7), weft count (x_8), warp count (x_9), air permeability (x_{10}), surface roughness (x_{11}), summit density (x_{12}), porosity (x_{13}), total surface of fibers (x_{14}) Plasma parameters Electrical power (x_{15}), treatment speed (x_{16})
Fabric surface wetting properties (outputs)	Cosine of the water contact angle (y_1), capillarity height (y_2)

Table 2 Selection of input variables relevant to water contact angle cosine

	Remaining inputs	Significance ranked by ascending order ΔS	Most relevant inputs	Irrelevant inputs
Step 1	All inputs, x_1 – x_{16}	$x_{15}, x_{16}, x_1, x_{12}, x_2, x_{11}, x_{10}, x_8, x_3, x_9, x_4, x_{13}, x_5, x_{14}, x_6, x_7$	x_{15}	x_7
Step 2	$x_1, x_2, x_3, x_4, x_5, x_6, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{16}$	$x_{16}, x_1, x_{10}, x_2, x_{12}, x_3, x_{11}, x_8, x_9, x_{13}, x_{14}, x_5, x_6, x_4$	x_{16}	x_4
Step 3	$x_1, x_2, x_3, x_5, x_6, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}$	$x_1, x_{10}, x_3, x_2, x_{12}, x_{11}, x_8, x_9, x_6, x_{14}, x_5$	x_1	x_5, x_{11}
Step 4	$x_2, x_3, x_6, x_8, x_9, x_{10}, x_{12}, x_{13}, x_{14}$	$x_{10}, x_3, x_2, x_{12}, x_{13}, x_8, x_9, x_{14}, x_6$	x_{10}	x_6, x_{13}
Step 5	$x_2, x_3, x_8, x_9, x_{12}, x_{14}$	$x_3, x_2, x_{12}, x_{14}, x_8, x_9$	x_3	x_9
Step 6	x_2, x_8, x_{12}, x_{14}	x_{12}, x_{14}, x_2, x_8	x_{12}	x_8
Step 7	x_2, x_{14}	x_2, x_{14}	x_2	x_{14}

According to Table 2, it could be noticed that electrical power (x_{15}), treatment speed (x_{16}), composition (x_1), air permeability (x_{10}), fiber count (x_3), weave construction (x_2), and summit density (x_{12}) are identified as the most relevant inputs for the cosine of water contact angle. The same result can be obtained for the capillarity height [10]. The only difference between them is that the orders of these two ranking lists of relevant inputs are slightly different. In this way, the number of input variables is reduced from 16 to 7. For both the cosine of water contact angle and capillarity height, the most relevant plasma process parameter is electrical power. The obtained two ranking lists are conforming to general professional knowledge of experts, and they also allow a better understanding on the plasma treatment process.

The networks models were based on a three-layer feed-forward neural network. The input layer corresponded to the seven selected inputs parameters. The output layer corresponded to the two outputs that were the cosine of the water contact

angle and the capillarity height. These networks were trained using two training algorithms: the Levenberg–Marquardt algorithm (LMNN model) and the Bayesian regularization algorithm (BRNN model). The performances of these networks were measured by computing the RMSE, the mean absolute errors (MAE), the mean relative absolute errors (MRAE), and the correlations coefficients (R) over the training and test data subsets. The LMNN model is optimized with 9 neurons in the hidden layer and 110 training iterations. The BRNN model is optimized with 7 neurons in the hidden layer and 250 training iterations. From these results, we can find that the Bayesian regularization algorithm is more efficient than the Levenberg–Marquardt algorithm because it has more reduced model structure and very little additional computational cost. The BRNN model is shown in Fig. 3.

Table 3 presents a comparison of the performances of the LMNN and BRNN models over the training and test data sets. It could be seen from this table that both the neural network models give high correlation coefficients and acceptable prediction errors, showing that their learning and generalization performances are good enough. Moreover, it can be noticed that the learning and prediction performances of the BRNN model are better than those of the LMNN model. Thus, it could be concluded that the Bayesian regularization method yielded higher prediction accuracy than the early stopping technique. Also, another advantage of this method is that it does not require any separate validation data set.

The relative importance of the input variables as calculated following the connection weight approach [14] is presented in Table 4. It shows that electrical power is the most important parameter for both the water contact angle cosine and the capillarity height. Besides, it is noticed that the ranking of inputs as per relative importance for both outputs is identical to that obtained from the fuzzy logic–based sensitivity variation criterion. It also shows that treatment speed, air permeability, and fiber count have a negative effect on both outputs. Conversely,

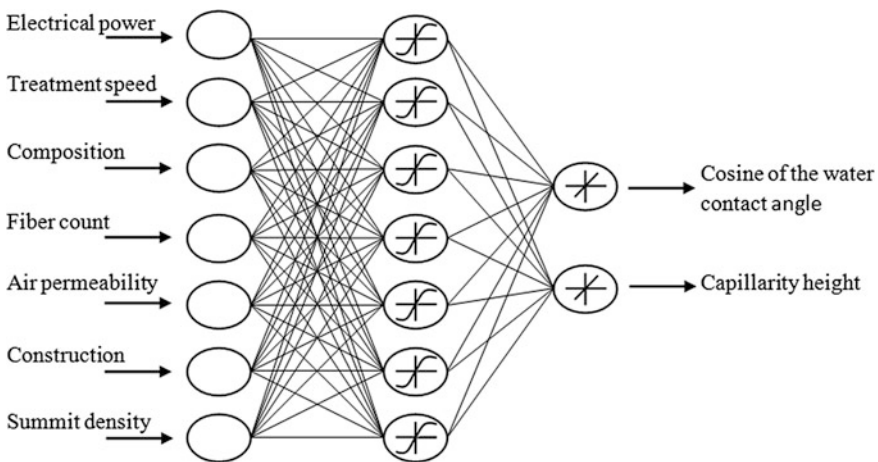


Fig. 3 BRNN model for water contact angle and capillarity height

Table 3 Comparison of the performances of the LMNN and BRNN models over the training and test subsets

		Training performance			
		RMSE	MAE	MRAE (%)	R
LMNN model	Water contact angle cosine	0.0156	0.0119	2.18	0.9952
	Capillarity height	1.0699 mm	0.8856 mm	4.82	0.9959
BRNN model	Water contact angle cosine	0.0107	0.0087	1.46	0.9977
	Capillarity height	0.8555 mm	0.6600 mm	3.92	0.9973
<i>Test performance</i>					
LMNN model	Water contact angle cosine	0.0208	0.0164	2.98	0.9920
	Capillarity height	1.2164 mm	0.9673 mm	5.08	0.9952
BRNN model	Water contact angle cosine	0.0137	0.0112	1.98	0.9957
	Capillarity height	1.0151 mm	0.7798 mm	4.02	0.9964

Table 4 Relative importance of different inputs as per connection weight approach for the prediction of water contact angle cosine and capillarity height by the BRNN

Inputs	Outputs			
	Water contact angle cosine		Capillarity height	
	S_i value as per connection weight approach	Ranking of inputs as per relative importance	S_i value as per connection weight approach	Ranking of inputs as per relative importance
Electrical power	4.2457	1	2.5842	1
Treatment speed	-1.5110	2	-0.9492	3
Composition	1.2493	3	1.2131	2
Air permeability	-0.7767	4	-0.8798	4
Fiber count	-0.4752	5	-0.4197	5
Construction	0.1114	7	0.2662	6
Summit density	0.1593	6	-0.1262	7

electrical power, composition, and construction have a positive effect on both outputs. Also, it can be seen from Table 4 that summit density has a positive effect on water contact angle cosine and a negative effect on capillarity height. In general, the results obtained from the proposed neural network model can effectively validate existing physical and chemical knowledge on the relationship between fabric structure and plasma treatment and help to generate new specialized knowledge in the related field.

6 Conclusion

In this paper, a feed-forward neural network model is used for understanding and predicting the relationship between plasma processing parameters and fabric surface wetting properties. The input variables of the model are selected using a fuzzy logic-based sensitivity variation criterion. Also, the proposed model optimization procedure has effectively improved its prediction accuracy. From our experiments, we find that the Bayesian regularization approach can lead to a better performance than the early stopping method on the training and test sets. The relative importance of the input variables is calculated using the connection weight approach. These results completely agree with those obtained using the fuzzy logic-based sensitivity criterion. Thus, it is believed that neural network models can be efficiently applied to understanding, evaluation and prediction of woven fabric surface modification by atmospheric air-plasma treatment.

References

1. Cai Z, Qiu Y, Zhang C, Hwang Y, McCord M (2003) Effect of atmospheric plasma treatment on desizing of PVA on cotton. *Text Res J* 73:670–674
2. Costa THC, Feitor MC, Alves C, Freire PB, de Bezzerra CM (2006) Effects of gas composition during plasma modification of polyester fabrics. *J Mater Process Technol* 173:40–43
3. Hossain MM, Herrmann AS, Hegemann D (2006) Plasma hydrophilization effect on different textile structures. *Plasma Process Polym* 3:299–307
4. Takke V, Behary N, Perwuelz A, Campagne C (2009) Studies on the atmospheric air-plasma treatment of PET (Polyethylene Terephthalate) woven fabrics: effect of process parameters and of ageing. *J Appl Polym Sci* 114:348–357
5. Coit DW, Jackson BT, Smith AE (1998) Static neural network process models: considerations and case studies. *Int J Prod Res* 36:2953–2967
6. Yousefi MMR, Mirmomeni M, Lucas C (2007) Input variables selection using mutual information for neuro fuzzy modeling with the application to time series forecasting. In: *Proceedings of international joint conference on neural networks, Orlando, Florida, USA*, pp 1121–1126
7. Fleuret F (2004) Fast binary feature selection with conditional mutual information. *J Mach Learn Res* 5:1531–1555
8. Thawonmas R, Abe S (1997) A novel approach to feature selection based on analysis of class regions. *IEEE Trans Syst Man Cybern B* 27:196–207
9. Deng X, Vroman P, Zeng X, Koehl L (2010) Selection of relevant variables for industrial process modelling combining experimental data sensitivity and human knowledge. *Eng Appl Artif Intell* 23:1368–1379
10. AbdJelil R, Zeng X, Koehl L, Perwuelz A (2009) Neural model of woven fabric surface modification by atmospheric air plasma. In: *Intelligent textiles and mass customization international conference, Casablanca, Morocco*
11. Huang YL, Edgar TF, Himmelbau DM, Trachtenberg I (1994) Constructing a reliable neural network model for a plasma etching process using limited experimental data. *IEEE Trans Semicond Manuf* 7:333–344

12. Gil P, Cardoso A, Palma L (2009) Estimating the number of hidden neurons in recurrent neural networks for nonlinear system identification. In: IEEE international symposium on industrial electronics. IEEE Press, Seoul, South Korea, pp 2053–2058
13. Lee KW, Lam HN (1995) Optimal sizing of feed forward neural networks: case studies. In: 2nd New Zealand two-stream international conference on artificial neural networks and expert systems. IEEE computer society, Dunedin, New Zealand, pp 79–82
14. Olden JD, Jackson DA (2002) Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks. *Ecol Model* 154:135–150

Model-Driven Approach for the Development of Web Application System

Jinkui Hou

Abstract In order to resolve the problems of model-driven software development, a model-driven approach for the development of Web application system is proposed in this paper. The development process starts from describing of platform-independent models. Then, mapping relations from the source model to the target model are built according to the syntactic structure and semantic features of both meta-models, and model transformation as well as code generation can be achieved subsequently. ASP.NET is used as a target platform in the experiment which shows that this approach follows the essence, process, and requirements of model-driven software development and thus can make an effect support for model-driven software engineering.

Keywords Model-driven software development • Modeling approach • Model transformation • Code generation

1 Introduction

High-level model description and model transformation are key technologies of model-driven architecture (MDA) [1], but there is no effective solution until now. Yang et al. [2] proposed a UML-based framework development approach, in which the concepts and semantics of UML are extended and throughout the entire software development process, but a clear description of software architecture is lacked. Meliá et al. [3] proposed that the architecture model of MDA could be extended with software architecture to improve the quality and efficiency of Web application development. However, they only gave an extension framework of

J. Hou (✉)

School of Computer Engineering, Weifang University, Weifang 261061, China
e-mail: jkhoul@163.com

MDA, and no specific model description approach and model transformation approach are presented. Convergence architecture proposed by Hubert [4] systematically describes architecture-centric software development, in which the process includes convergence business object modeling, convergence model refinement, UML refinement, and code generation. However, the architecture style of target platform is introduced at the initial stage of convergence architecture, which affects the independence of PIM platform. Similarly, the models built with some commercial MDA tools, such as OptimalJ, Rational XDE, Arcstyler, and AndroMDA, are not the platform-independent models in MDA sense.

Focus on the problems mentioned above, platform-independent Web application models are established under the guidance of software architecture from the view of implementation of software engineering. Then, semantic features of the source model are reconstructed in the target semantic domain. The mapping rules between models are defined according to grammatical structure and semantic characteristics of modeling elements in both ends. Lastly, the target system is achieved through a series of model transformation and software development is completed.

2 Platform-independent Modeling Approach

The UML-based modeling approach ASLP [5] is used to build source model, and ASP.NET is chosen as the target platform in the experimental research.

The ASLP approach is based on extending UML and adds user-interface presentation views. In this approach, there is an abstract description of UI component data and behavior elements rather than a list of interface elements and its attributes. The binding relations between UI elements and the corresponding objects are given at the same time, which made UI elements being platform-independent and the data elements as well as behavior elements are independent to some specific UI elements. ASLP can be used to create platform-independent models for Web applications as the source in model transformation, and it was composed of two levels: architecture modeling and component modeling.

In architecture models of ALSP, system represents the architecture and constraints of a software system, which is defined as a 4-tuple: $\langle S, D, C, R \rangle$. Herein, S represents architecture style. D represents function description for the system. C represents the set of components and connectors. R is a list of relations among components and connectors. Component is composed of component ports and its internal realization modules. Connector is a special kind of component used for changing information and maintaining action links.

After architecture design is completed, a realization component is needed for each component model. Existing components are expected to reuse, and non-existent components should be developed. Under the constraints of architecture model, component and connector should be described using UML according to the system's evolutionary requirements.

Table 1 The mapping relations from ADL to UML

ADL	UML
Component	Package
Complex connector	Package
Functional specification	Interface
Entry point	Abstract class

Component and complex connector are mapped to package, and port is mapped to abstract class. Functional specification is mapped to interface. The association relations between components (or between component and complex connector) are mapped to the relations between classes (the access point of the caller and the class that implements the interface of the callee). The mapping relations from architecture model to UML are shown in Table 1.

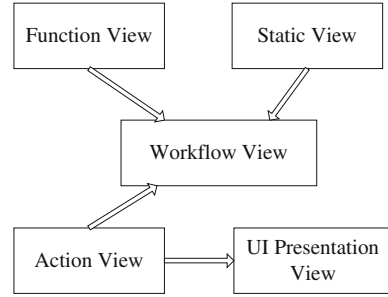
As shown in Fig. 1, function view, workflow view, static view, action view, and UI presentation view are used in ALSP approach for building component models. Each view represents an aspect of the application system.

The functions of a component in architecture model exchange information between the system and outside, and the interactions among function modules of the system are all described using function views. It uses use-case diagram in UML to complete description. Workflow views are used to modeling the actions of each individual and define interactive relations and cooperative relations among these entities. It uses state-machine-based activity diagram to complete description. Static view is an integration of package diagram and class diagram in UML. It is used to describe analytical classes of the use cases in function view and the relations among these classes. Action view uses extended collaborative diagram in UML to describe the actions of objects in more detail. UI presentation view provides intuitional presentation for boundary objects and the interaction points between user and system in action view. It also provides binding relations between UI modeling elements and the visible elements in action view.

3 Model Transformation and Code Generation

ASP.NET [6] is a framework for the development of Web applications, and C# is used as the target language in the experiment of this paper. The macrostructure of the target application model based on ASP.NET includes: (1) Web pages (.aspx), and it contains UI elements and the foundation codes; (2) the code-behind file (.cs), which is used to achieve the logical separation of Web pages code and of user interface; (3) web.config, which are XML-based configuration files; (4) global.asax, which is the global application file, you can define global variables and global react to events; (5) bin directory to store application to use. NET assemblies and the compiled code; and (6) Web project file, compiled code and optional Web services (.asmx) and user controls (.ascx).

Fig. 1 Relations among the views of component model



3.1 Model Mapping Relations

The semantics mapping between models can be considered as a constructing process in the target semantic domain for source semantic domain. That is, starting from source models, the values of relevant attributes can be obtained through observation and deduction on the source elements, and then to ascertain whether these values meet the requirements for the definition of target models.

Let C be the set of concepts of patterns in the target model, $C = \{c_1, c_2, \dots, c_n\}$; Let O_1 be the set of attributes which can be observed directly, and O_2 be the set of attributes observed implicitly, that is to say, these attributes of source models need to be deduced by using the context of concepts in the pattern, that is, the necessary condition of these concepts, which noted as $\{N_c | c \in C\}$; Let R be the classification rules for the attributes values of target semantic domain, that is, the sufficient condition for the classification in target semantic domain, which noted as $\{S_c | c \in C\}$. Let M is the mapping relations between patterns. The mapping process is to find a conceptual set C_i in target semantic domain for each conceptual element C_s in source domain, which satisfy the following equation.

$$O_1 \wedge O_2 \wedge R \Rightarrow M(c_s, c_i) \quad (1)$$

The equation given above describes the mapping problem in formalism, that is, a set of concepts of pattern are given, which noted as $CS = \{c_1^S, \dots, c_m^S\}$. Then, it can be mapped to target domain using the classification rules for the attributes values in target semantic domain noted as a concept set $CT = \{c_1^T, \dots, c_n^T\}$. The mapping process is depended on semantics features in both ends. The source provides the observation of the source pattern ($O = \{N_c | c \in CS\}$). The target provides target pattern and its classification rules ($C = CT, R = \{S_c | c \in CT\}$). By this way, a pattern conceptual element can be mapped into target semantic domain through the process of finding the target conceptual set c_i^T , which satisfy Eq. (1).

According to the above principles, mapping relations from the source model to the target model are built according to the syntactic structure and semantic features of both meta-models, and model transformation can be achieved subsequently. The

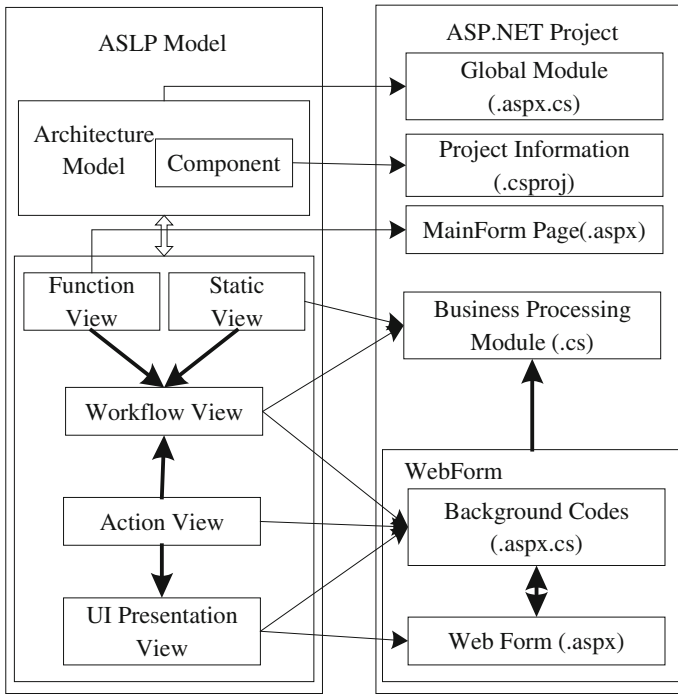


Fig. 2 The mapping relations from source model to target model

mapping relations from the source model based on ASLP to ASP.NET project model are shown in Fig. 2.

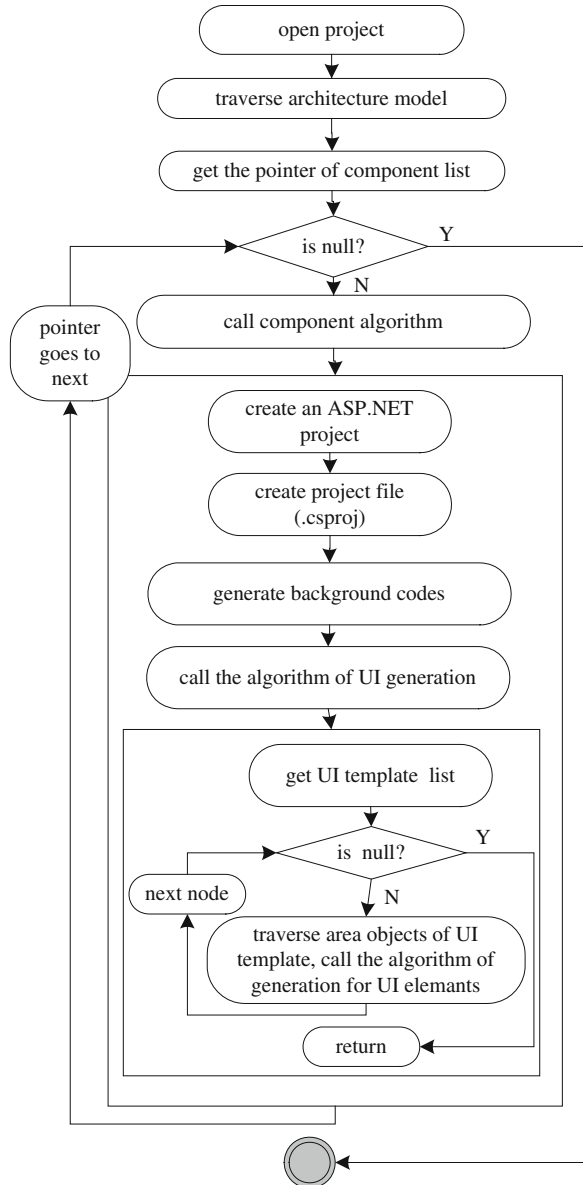
The information of component configuration and association is mapped into global files, global variables, and public functions. Project components of architecture model is mapped to ASP.NET project, and engineering information of components combined with information about other models is mapped into engineering information of the project (*.csproj file). Object view is mapped to the business processing module of the generated ASP.NET project, which provides corresponding support for interaction view and presentation view. Compound use case of function view is mapped to functional selection for the project. Module processing information of presentation view and interaction view are mapped to the background processing codes of Web pages generated by presentation view (*.aspx.cs files). Use cases of interaction view are mapped to the corresponding operations for the menu, button, or hyperlink of UI. Relations of method invocation are mapped to call relations of the corresponding method. The UI navigation relations are mapped to the display operations for the target Web pages.

Template object of presentation view is mapped to Web page, in which the information is mapped to the type of elements, location, size, color, and other attribute information of Web form (*.aspx files).

3.2 Code Generation

Target code generation algorithm can be summarized as four major functions: generation of framework of target codes, generation of component codes, code generation for user interface, and code generation for interface elements. The process is shown in Fig. 3.

Fig. 3 The flow of the algorithm for target code generation



The algorithm for generation of framework of target codes is the entrance of the whole automatic code generation. The objects of architecture models are traversed in loop, and different target project is generated by the appropriate code generation algorithm according to the type of components.

The functions of the algorithm for component code generation mainly include building an ASP.NET project; generating the main module of the project and the documents required by other services according to architecture model; initializing the main form of target project; traversing the interface template and invoking interface generation algorithms.

The algorithm of code generation for user interface generates Web page codes in terms of the information provided by presentation models (interface templates). The elements include window frame, interface elements, and UI layout. At the same time, the corresponding background operation codes are generated according to the information of the interaction view and object view corresponding to the UI presentation view. The main process of the algorithm is as follows: traverse the area objects of the interface template, then invoke the corresponding code generation algorithm according to the type and display form of data of each interface presentation unit.

The generation algorithm for UI elements is relatively complicated. The objects of UI template cannot be mapped to the specific controls in the ASP.NET environment one by one. Thereby, the presentation objects of UI template are mapped to corresponding ASP.NET controls according to the display style. If the control contains child controls inside, it is necessary to generate in a recursive way.

We have basically realized prototype supporting tool, which main functions include (1) graphical architecture modeling; (2) mapping from architecture model to OOD design model; (3) graphical component modeling; (4) transformation from high model to the target platform model; (5) management of component library; (6) model validation; and (7) automatic target code generation. All programs are realized in the MicrosoftTM Visual C++.NET.

4 Conclusion and Future Work

A Web application development approach is proposed in this paper on the basis of combining software architecture and MDA. The platform-independent features of high-level models are maintained. Software development is achieved through automatic transformation between models. This approach follows the essence, process, and requirements of model-driven software development and thus could make an effect support for model-driven software engineering. The main future work is the formal verification of model transformation and the further detailed description of the user interface.

Acknowledgments The author is most grateful to the anonymous referees for their constructive and helpful comments on the earlier version of the manuscript that helped to improve the

presentation of the paper considerably. This research was supported by the foundation of science-technology development project of Shandong Province of China under Grant No. 2011YD01042 and No. 2011YD01043.

References

1. Miller J, Mukerji J (2011) MDA guide version 1.0.1 (document number omg/20011-06-01). <http://www.omg.com/mda>
2. Yang YJ, Kim SY, Choi GJ et al (2008) A UML-based object-oriented framework development methodology. In: Proceedings of Asia Pacific software engineering conference 2008, Taipei Taiwan. IEEE Press, New York, pp 30–39
3. Meliá S, Cachero C, Gómez J (2010) Using MDA in web software architectures. In: Proceedings of 2nd international workshop on generative techniques in the context of MDA, Anaheim, California, USA. IEEE Press, New York, pp 76–82
4. Hubert R (2002) Convergent architecture: building model-driven J2EE Systems with UML. Wiley, New York
5. Hou J, Wan J, Yang X (2006) MDA-based modeling and transformation approach for WEB applications. In: Proceedings of the sixth international conference on intelligent system design and applications (ISDA). IEEE Computer Society, New York, pp 867–812
6. Jeffrey R, Francesco B (2009) Applied Microsoft.NET framework programming. Microsoft Press, Washington

A Web UI Modeling Approach Supporting Model-Driven Software Development

Jinkui Hou

Abstract Model-driven software development has become a tendency in software engineering. However, Web user interfaces have the characteristics of the customization but frequently renewing, which makes traditional software development approach not suitable for the design requirements of Web pages. To solve the problems of the development of Web user interface, and focusing on the characteristics of Web application, a user interface modeling approach is proposed on the basis of interface template and XML technology. Web user interfaces are described in a direct-viewing style with graphics at the model level. The approach can provide an effective support for model-driven software development.

Keywords Web user interface · Interface template · XML · Model-driven development

1 Introduction

With the rapid development in network technology, the requirements of the user continue to increase for the development efficiency and quality of Web applications, which leads to the increasing difficulty in development. Model-driven software development is becoming a trend in software engineering, in which user interface modeling is an important part. The core idea of the traditional conceptual model of user interface, such as PAC model and MVC model, is the separation between display and logic functions, but it is not well-supported model-driven software development. General interface design tools are simple and easy to use, but they only describe the static interface and do not support interaction. With the

J. Hou (✉)

School of Computer Engineering, Weifang University, Weifang 261061, China
e-mail: jkhoul@163.com

progressive development in model-driven software development and the growing separation between software functions and interface presentation, it has become the trend of software design to generate the interface according to the interface model [1, 2]. There are many model-based methods and tools, such as UIDE [1], MASTERMIND [2], UMLi [3], but they can only provide some design descriptions and recommendations and not provide effective support to the final interface generation and UI layout. Furthermore, there is no detailed consideration on the special requirements of the user interface modeling in Web environment.

Web interface is centered by information display, and the user is a complexity. The requirement of Web UI not only includes high availability and robustness, but also includes more visual appeal. It plays a critical role for the layout and presentation form of user interface to the success of the whole software [3]. A good UI presentation model must be extracted from different domains, so as to provide a better guide to the automatic generation of Web user interface.

To solve the problems mentioned above, a Web user interface modeling approach is proposed on the basis of interface template and XML technology, which can provide an effective support for model-driven software development.

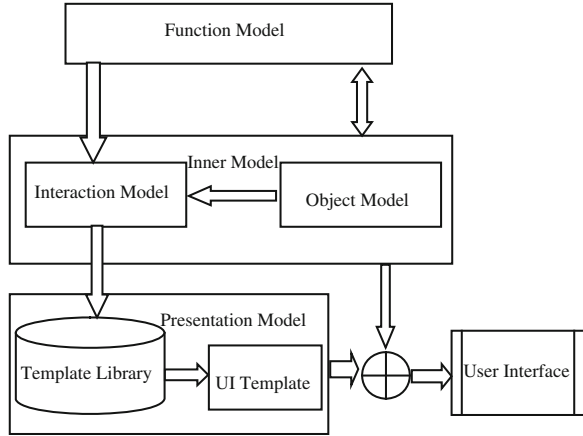
2 FMP Model

The function, model, presentation (FMP) [4] modeling approach is on the basis of the traditional application modeling approach by adding a description view of user interface, which is shown in Fig. 1. The model of this approach is not a description of the specific forms of interface elements and their attributes, but the description of the abstract data and behavioral elements of user interface. At the same time, the corresponding relationships between interface elements and presentation objects are also shown. Thereby, the interface elements are independent with any specific application platforms, and the data elements, and behavior elements, are separated from specific interface elements. As the source of model-driven development, the FMP modeling approach can be used to build platform-independent Web application model.

The extended use case diagram in UML is used in function model, which determines the UI needs to the internal model and the relationship between various UI pages through the analysis of the requirements of users.

Internal model includes object model and interaction model, in which object model describes the system from the static view, and it describes the necessary objects that constitute UI interface and the relationship between them. Interaction model describes the system from the dynamic view, which describes the composition of UI and the relationship between interface elements on the basis of object model and function model. The abstract form of the user interface can be fully expressed through the internal model. Presentation model shows the layout and presentation style according to the internal model and the display requirement of data from the users. It is a full description of the user interface in direct-viewing

Fig. 1 The structure of FMP model



style, so that the interface elements are independent on any specific application platform, and data elements and behavioral elements are separated from the specific interface elements.

3 User Interface Template

Rich experience has been accumulated in the process of software development. In this paper, the FMP model is extended in order to guide later software development. The organizational structure of interface templates and template libraries is improved, which can provide an intuitive presentation at the modeling stage for the previous development experience and interface design patterns. Thereby, the efficiency and quality of software development are increased.

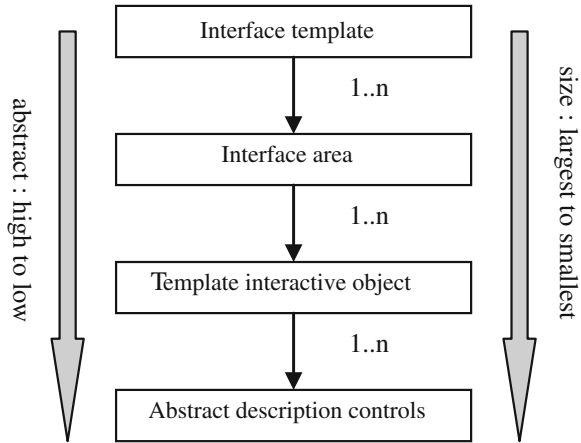
3.1 The Structure of Interface Template

As shown in Fig. 2, the presentation view is composed by interface templates, interface area, template interactive object, and abstract description controls.

An interface template is divided into multiple interface areas according to the layout, while a template interactive object consists of one or more abstract description controls. A template interactive object is composed of one or more abstract description controls.

At interface template layer, the interface style and macrolayout should be determined. In interface regional layer, main consideration should be given to the display form of the interactive object. In the abstract description controls, layer's main consideration should be given to intuitive interface presentation.

Fig. 2 The structure of UI presentation model



The four-layer structure of interface design makes it easier to control parameters of the template, which separates concerns at different abstract levels.

3.2 The Layout Tree of Interface Template

The layout tree of interface template describes the interface layout strategy. A Web page is divided into different zones. There are two adjacent relationships between these zones: horizontal adjacent and vertical adjacent.

As shown in Fig. 3, the whole page is divided into left and right parts through vertical division.

The left part is a tree style, and the right part is divided into upper and lower sections through horizontal division. The upper part is the graphics and the lower part, a table.

The leaf nodes of the layout tree represent abstract description components, and the non-leaf nodes represent the division mode of the interface template. According to this layout style, the layout of the structure tree can be easily described and its XML representation is shown as in Fig. 4.

4 Library of Interface Templates

According to the classification criterion of interaction model, a variety of templates and common abstract description widgets are stored in the library of interface templates in the form of XML files. The template library consists of three parts: storage subsystem of template library, query subsystem, and maintenance subsystem. Storage subsystem can be a professional relational database and also

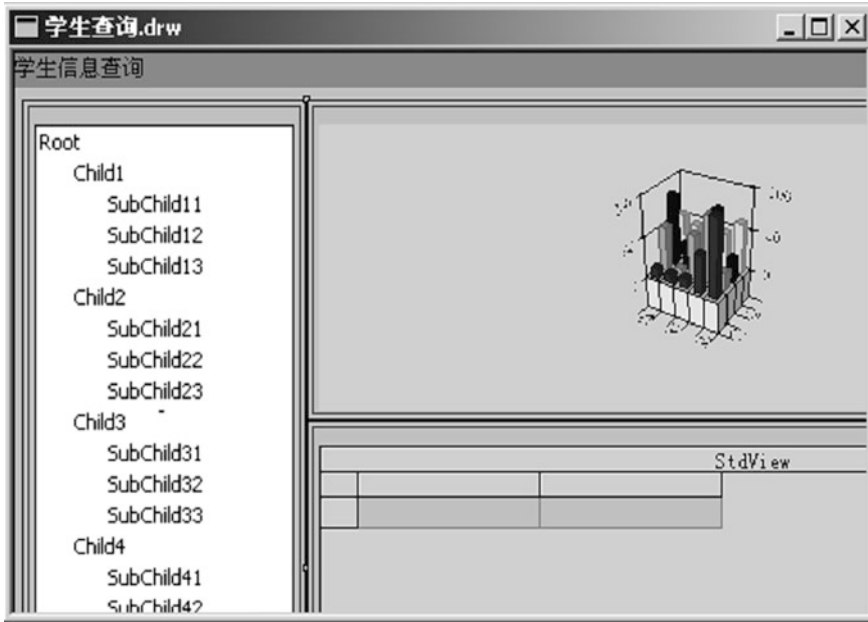
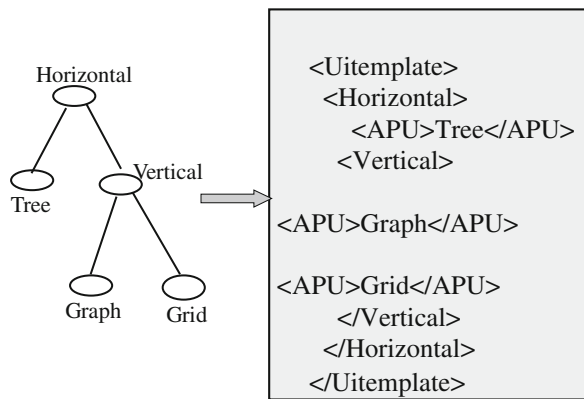


Fig. 3 An instance of interface template

Fig. 4 Layout tree and its XML representation



can be a series of XML files. Query subsystem is responsible for the communication with interaction model and interface template, which provides the support for query. The maintenance subsystem is responsible for the adding, editing, and deleting UI templates in the template library.

4.1 The Composition of Interface Template Library

In the process of constructing the interface template library, the first is the analysis of the appearance, characteristics, and behavior of the overall interface and its elements and analysis. Then, we extract their common features and form general interface templates and abstract description model. They are divided in accordance with the types and levels. First of all, these elements are classified by types, and then, the common characteristics of the same type of elements are taken out and are putted on the upper layer, and so on. Finally, this approach comes down to the root node.

The basic elements of Web interface can be abstracted into five categories: data objects, data collection, querying object, object for control parameters, and object for use case collection. Data object is mainly used for data input and output. Different attributes of the object use different types of control components to display, and this presentation format is called free form. The properties belonging to the same group are located within the same framework to improve the readability of the interface logic. Data collection is a presentation form which is used for concentrated displaying and operating a collection of data objects, which mainly includes free form, tabular form, tree form, graphic form, and the form of object groups. Querying object is used to generate a query condition. When the querying object is a data object, its presentation form is usually a simple control group, which is generally associated with buttons or hyperlinks. When the querying object is a data collection, its presentation style is tables or trees. Control parameter object is used to generate the control parameters, and its presentation forms usually are drop-down list box, checkbox, and other simple control group. The use cases represent related operations, and their presentation form can be ordinary buttons, image buttons, menu items, hyperlinks, and so on. Use case group is a collection of use cases applied to the same data object, of which general forms can be ordinary button group, picture button group, and hyperlink group.

4.2 Matching of Interface Template

In the interface template library, there are many interface templates with similar functions. The one with the highest degree of matching should be filtrated for the user. According to the component querying methods proposed in [5], we take two steps to retrieve the best one.

Firstly, according to the number and type of abstract description controls and of the overall style of the template, the overall framework of the interface template is determined, and a set of eligible templates is identified.

Then, the highest matching interface template is determined by defining the degree of structure matching (SM) and the degree of attribute matching (AM). Herein, SM represents the matching degree of the similar templates in the aspects

of layout, etc. AM refers to the matching degree of the similar templates in the aspects of overall properties, associated semantics, etc.

The weight of all leaf nodes of the matching model (noted as MV) is defined as 1, and the weight of the parent node is the sum of the weight of all its child nodes. QR and TR, respectively, represent the structure of the required template and the template structure in the database. QA and TA, respectively, represent the attributes of the required template and the template attributes in the database. $M(QR,TR)$ represents the matching degree between the querying condition and the atomic structure of templates in database, while $M(QA,TA)$ represents the matching degree between the querying condition and the properties of templates in database. The value is set to 1 if match and 0 if not.

The formula for SM degree is shown as follows:

$$SM = \frac{\sum_{i=1}^n M(QR_i, TR_i) * MV_i}{\text{Number of (TR)}} \quad (1)$$

The formula for AM degree is shown as follows:

$$AM = \frac{\sum_{j=1}^m M(QA_j, TA_j) * MV_j}{\text{Number of (TA)}} \quad (2)$$

Thereby, the formula for the whole template matching degree is shown as follows:

$$\text{Match}(T) = SM + AM \quad (3)$$

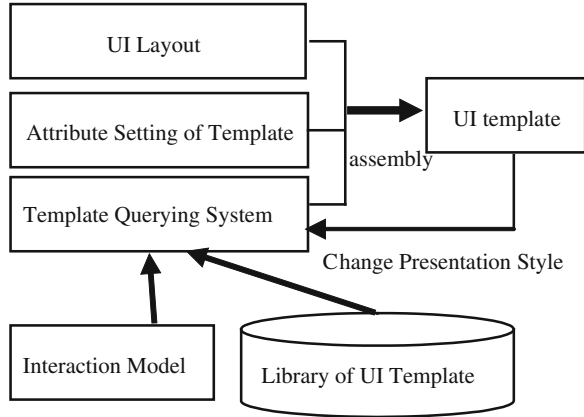
Based on the above formulae, the most qualified template can be calculated relatively easily.

5 Verifying with Code Generator

Code generator generates a specific Web page according to the layout information, presentation style, and description parameters of interface template. The detailed process is shown in Fig. 5.

First of all, presentation model gets the name of an abstract description control widget from interaction model, and then, it is placed into the template query subsystem; On the basis of input information, the system begins to query in the library of interface templates. The result will be stored in the form of XML text block. At the same time, the user creates a new interface template and sets the parameter and layout according to the user's requirements. The parameters and layout set are saved in the XML file describing the interface template. And then, the XML text chunk inquired is integrated into the template XML file according to

Fig. 5 Assembly process of UI template



the layout mode, and the abstract description control is displayed in the template. And the user modifies the relevant parameters and presence according to the circumstances of each case. Parameters modified are saved in XML files describing the interface template.

In addition, users also can directly match the eligible common template in the template library according to the number and type of abstract interaction control widgets from interaction model, thus without having to manually set the layout and property.

The approach allows the user to reuse the interface in the process of the modification. That is to say, the user can choose a ready-defined template according to the situation and insert it into a zone of another template, or copy an existing template when creating the new interface template in order to improve the reusability of interface template and the degree of automatic code generation.

In code generation, XML files of the template will be passed to the code generator to provide the required parameters for generating Web interfaces. The finally generated codes are separated from the model. If the generated user interface is not satisfied, it can be redesigned by modifying the parameters of the XML files or by directly adjusting the model in the model editing environment.

Model designer can use a professional XML editor to describe the new controls or the interface template and store these new abstract description controls and interface templates into template library according to their classification. By using editing tool, the designer can preview the abstract description controls and UI template, thus to make sure if they meets the requirements.

We have designed a Web application development system based on FMP, all of the programs are achieved within the environment of Microsoft™ Visual C++.NET. At present, automatic code generation has been completed from high-level model to J2EE and ASP.NET. Figure 6 shows an example of interface template of adding employees in a human resource management system, of which the display area is formed by the three subregions. The running page of the generated Java codes on the Struts framework is shown in Fig. 7.

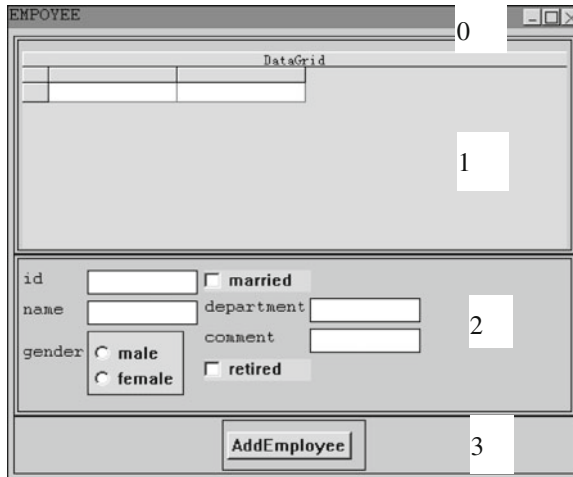


Fig. 6 An example of interface template—adding employees

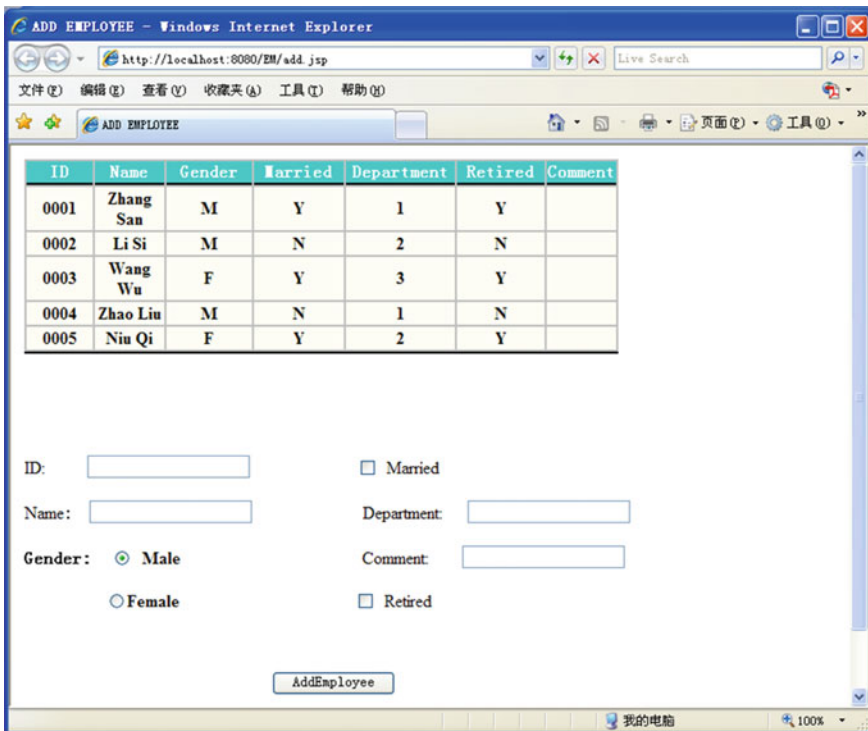


Fig. 7 A running instances—adding employees

It can be seen from this example that the approach can provide supports for automatic generation of Web user interfaces.

6 Conclusion and Future Work

In this paper, a user interface modeling approach is provided on the basis of XML and interface template, and the reusing of interface design patterns is achieved. The approach can provide an effective support for model-driven software engineering [6]. As far as our future work is concerned, a comprehensive extraction and description of Web interface model will be conducted to further enhance the visual appeal of the generated Web pages.

Acknowledgments The author is most grateful to the anonymous referees for their constructive and helpful comments on the earlier version of the manuscript that helped to improve the presentation of the paper considerably. This research was supported by the foundation of science–technology development project of Shandong Province of China under Grant No. 2011YD01042 and Grant No. 2011YD01043.

References

1. Won K, Don JF (2010) User interface presentation design assistant. In: Hudson S (ed) Proceedings of UIST'10, New York. ACM Press, New York, pp 10–20
2. Pedro S, Piyawadee S, Palbo C et al (2008) Declarative interface models for user interface construction tools: the MASTERMIND approach. In: Proceedings of engineering for human–computer interaction. Chapman & Hall, London. IEEE Press, New York, pp 120–150
3. Silvapp DA, Paton NW (2009) User interface modeling in UMLi. *IEEE Softw* 20(4):62–69
4. Wan J, Sun B (2011) Interface model to support automated generation of user interface. *Comput Eng Appl* 39(18):730–736
5. Yan L, Feng Y, Song Y (2010) Reusable component classification and querying method. *Comput Eng Appl* 39(6):85–87
6. Kleppe A, Warmer J, Bast W (2009) MDA explained, the model driven architecture: practice and promise. Addison-Wesley, Boston

On the Possibilistic Handling of Priorities in Access Control Models

Salem Benferhat, Khalid Bouriche and Mohamed Ouzarf

Abstract Access control models are important tools for modelling security policies. They allow to limit the access to sensitive data to only authorized users. This paper focuses on organization-based access control (OrBAC) model which represents a generic framework for compactly representing general security policies rules. More precisely, we propose to add to OrBAC model a new entity, called priority, that encodes different forms of uncertainty that may be encountered in security rules. These priorities will be modelled in possibility theory which represents a natural framework for handling uncertain information. We propose different combination rules that allow to derive concrete permissions from prioritized abstract permissions.

Keywords Possibility theory · Possibilistic logic · OrBAC

1 Introduction

In many applications, it is important to protect sensitive data from unauthorized users. Access control models are appropriate solutions to solve this problem by limiting access to data on the basics of security policies. Different access control models have been proposed in literature: Harrison, Ruzzo, and Ullmann (HRU)

S. Benferhat (✉) · K. Bouriche
CRIL-CNRS, Centre de recherche en informatique de Lens, Université d'Artois,
Rue Jean Souvraz, 62307 Lens Cedex, France
e-mail: benferhat@cril.fr

K. Bouriche
e-mail: bourichekhalid@hotmail.com

K. Bouriche · M. Ouzarf
Département de l'informatique FST de Fez, University Sidi Mohamed Ben Abdellah
(USMBA), Fès, Maroc
e-mail: ouzarf@yahoo.com

[1], discretionary access control (DAC) model [2], team-based access control (TBAC) model [3], role-based access control (RBAC) [4–6], organization-based access control (OrBAC) [7], etc.

This paper proposes to integrate an uncertainty component in access control models. In practice, the assignment of a permission, to achieve some action on an information system, is not always fully specified. In addition to contexts, a permission may be pervaded with uncertainty due for instance to the presence of exceptions or simply due to the reliability of the source. For instance, one may provide some law rules that may justify some permissions (or prohibitions). However, it may happen that some doubt/uncertainty on such certain law rules may exist. For instance, a law rules may be an old one, and there is chance that more recent (and priority) laws, with a contradictory recommendation, have been proposed. Uncertainty here is more represented by some ordering relation between security rules and is rarely described by probability distributions.

Priorities or uncertainties considered in this paper are represented within a possibility theory framework. Possibility theory and possibilistic logic [6–8] are uncertainty theories that allow to deal with uncertain information in an ordinal way, namely only the priorities between information is important. Possibilistic logic is a simple extension of propositional logic, by assigning a weight to each formula. This weight represents the uncertainty/priority of the formula. The highest is this priority, the most important/proprietary is the formula.

The access control model considered in this paper is the one of organization-based access control (OrBAC) model [7]. The choice of OrBAC is mainly justified by its expressive power to represent general security rules. The main idea is to separate general knowledge, representing generic rules, from factual data representing concrete elements of an information system. In OrBAC, the relations between concrete elements of the system and their abstraction are often provided by crisp relations that do not admit uncertainties or exceptions. This paper shows how to extend these relation entities, to deal with uncertainty. This allows to enhance OrBAC systems with a new feature that integrates priorities entities in different relations of OrBAC system.

The rest of this paper is organized as follows. [Section 2](#) gives a brief refresher on possibilistic logic and possibility theory. [Section 3](#) provides a brief overview on existing access control models. In [Sect. 4](#), we introduce our extension of OrBAC model. In [Sect. 5](#), we provide different combination modes of priorities to derive concrete permissions from prioritized abstract permissions. [Section 6](#) concludes the paper.

2 A Brief Refresher on Possibilistic Logic

Information is often pervaded with uncertainty, and logics of different types have been developed for handling uncertain pieces of knowledge. Possibilistic logic [8] offers a convenient tool for handling uncertain or prioritized formulas and coping

with inconsistency. Propositional logic formulas are thus associated with weights belonging to a linearly ordered scale.

Possibilistic knowledge bases are made of a finite set of weighted formulas:

$$\Sigma = \{(\phi_i, a_i) : i = 1, n\}. \quad (1)$$

where a_i is understood as a lower bound of the degree of necessity $N(\phi_i)$ (namely $N(\phi_i) \geq a_i$). Formulas with null degree are not explicitly represented in the knowledge base (only beliefs which are somewhat accepted by the agent are explicitly represented). The higher the weight, the more certain the formula.

A possibilistic knowledge base Σ is said to consistent if the classical knowledge base, obtained by forgetting the weights, is classically consistent.

Given a possibilistic belief base Σ , we can generate a possibility distribution from Σ by associating with each interpretation, its level of compatibility with agent's beliefs, namely with Σ , as explained now.

A possibility distribution π is a function from the set of interpretations to $[0, 1]$. It extends the concepts of models and countermodels of propositional logic.

Therefore, the possibility distribution associated with $\Sigma = \{(\phi, a)\}$ is

$$\forall \omega \in \Omega, \pi_{\{(\phi, a)\}}(\omega) = \begin{cases} 1 & \text{if } \omega \vdash \phi \\ 1 - a & \\ \text{otherwise} & \end{cases} \quad (2)$$

When $\Sigma = \{(\phi_i, a_i), i = 1, n\}$ is a general possibilistic belief base, all the interpretation satisfying all the beliefs in Σ will have the highest possibility degree, namely 1, and the other interpretations will be ranked w.r.t. the highest belief that they falsify, namely we get [8]:

$$\forall \omega \in \Omega, \pi_{\{(\phi, a)\}}(\omega) = \begin{cases} 1 & \text{if } \forall (\phi_i, a_i) \in \Sigma, \omega \vdash \phi \\ 1 - \max \left\{ a_i : (\phi_i, a_i) \in \Sigma \text{ and } \omega \not\vdash \phi \right\} & \\ \text{otherwise} & \end{cases} \quad (3)$$

3 A Brief Overview on Access Control Model

Access control is an important area in information security. It provides tools to restrict access to sensitive information. There are different access control models that have been proposed in the literature: HRU [1], RBAC [4–6], TMAC [9], DAC [2], mandatory access control (MAC) [10, 11], team-based access control (TBAC) [12], etc.

3.1 Lampson and HRU Models

The Lampson model [2] introduced matrix access control in 1971. It structured the model as a machine state, where the triple (S, O, M) represents each state, the triplet means:

- S is a set of subjects
- O is a set of objects
- M a matrix of access control

Lampson's model associated with DAC does not directly express prohibitions or obligations. A user gets a permission to run some action on an object, if such action is explicitly stated in the matrix access.

HRU [1] also uses an access matrix. It was designed to improve the Lampson model and solve the problem of the complexity to update the access matrix by specifying commands which may be applied. These commands are used for updating the matrix and security policies. When new subjects, new objects, or new actions are introduced in the system, it is necessary to update the security policy in order to record the permissions granted to these new entities. One of the limits of the HRU model is that it only enables the administrator to specify permissions. Neither prohibitions, obligations nor recommendations are included in HRU models.

3.2 DAC Models

The main idea in discretionary access control is that handling access rights to each object is in the absolute discretion of the owner or the subject who is responsible. In DAC [2] model, any subject who creates an object becomes his owner, hence he may grant rights to other subjects, and then he can bring the security policy in an inconsistent state. This model is usually implemented with access control lists (ACL). A simple example of DAC model is the one used by UNIX operating system.

The strength of the discretionary policy is based on complete trust given to subjects that run the accounts of users and users themselves. Such models are not appropriate for sensitive applications such as the ones of medical area.

3.3 RBAC Models

The RBAC [6] defines an authorization model for accessing resources. Its main objective is to simplify security management by introducing the concept of role which is a factorization of assigning permissions in order to prevent more relationships between permissions and users.

The role represents an abstraction of an activity. RBAC was enriched by other variants, including the concepts of session, hierarchy of roles, and constraints on the roles.

4 Adding Priorities to OrBAC Model

This section focuses on organization-based access control (OrBAC) model proposed in Ref. [7]. This model is centered on the concept of organization and modelled in first-order logic. OrBAC offers a possibility to take into account the notion of contexts, which enables to represent conditional rules. OrBAC also allows to model the three deontic modalities: *permission*, *prohibition*, and *obligation*. Besides, different extensions have been proposed to deal for instances with temporal contexts [13] or to handle the administration of security policies [14].

This section consists in proposing another extension of OrBAC models. The idea is to integrate uncertainty in different components of OrBAC models. We review these basic concepts of the OrBAC model [7], using a diagrammatic language based on the entity relationship model. For each component, we integrate the *priority* entity to reflect the uncertainty that may present in defining security rules.

4.1 Organizations

An important entity in the OrBAC model is the entity *Organization*. An organization can be seen as an organized group of subjects playing same roles. As we will see below, all main concepts of security policies are defined within an organization.

4.2 Subjects and Roles

In OrBAC model, a *Subject* can be either a user (e.g., Tom or a process) or an organization (e.g., Linux distribution).

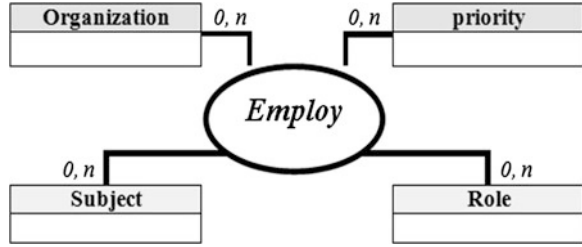
The entity *Role* is used to structure subjects in organizations. Roles associate subjects that fulfill the same functions.

Seeing that subjects play roles in organizations. We need a relationship that joins up these entities together. This relationship may be uncertain. This typically happens when there is inheritance between roles. In some situations, some inheritance may be questionable.

The relationship *Employ* (org, s, r, p) (see Fig. 1) means that the organization org employs subject s in role r .

p reflects the uncertainty in the relation *Employ*. It is a real number that belongs to the interval $[0, 1]$, where $p = 1$ means that the subject s definitely plays the role r in the organization, while p close to “0” means that there is a serious double in

Fig. 1 Employ relationship



the *Employ* relation. Note that contrarily to possibilistic logic, p is included in the parameter of the predicate. Here, formulas are not pairs (φ, α) as in standard possibilistic logic but first-ordered formulas where α is an argument of predicates. This is done for sake of simplicity in order to remain within entities relationships representation.

4.3 Objects and Views

In the OrBAC model, the entity *Object* covers inactive entities. In Linux operating system, an example of objects is the “ordinary files.”

The entity *View* corresponds to a set of objects that satisfy a common property. Seeing that views characterize the ways that objects are used in organizations, the relationship *Use* (org, o, v, p) means that the organization org uses object o in view v .

p again reflects the uncertainty associated in viewing the object o as the view v . Here, p may reflect a typicality relation. For instance, consider the Linux concrete action “grep.” Consider two activities “reading documents” and “searching expressions in documents.” Clearly, “grep” is a typical action of “searching expressions in documents,” while “grep” may be used in activity “reading documents” but it is not a typical action for such reading activity (Fig. 2).

4.4 Actions and Activities

In the OrBAC model, the entity *Action* contains computer actions such as “read,” “write.” Seeing that subjects and objects are abstracted by means of roles and

Fig. 2 Use relationship

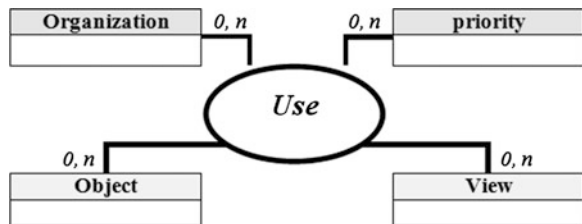


Fig. 3 Consider relationship

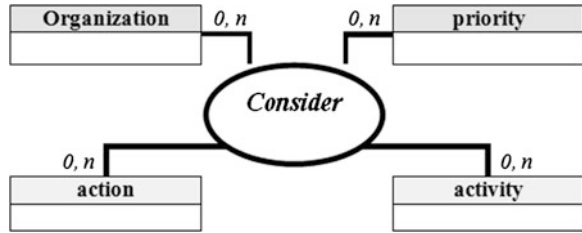
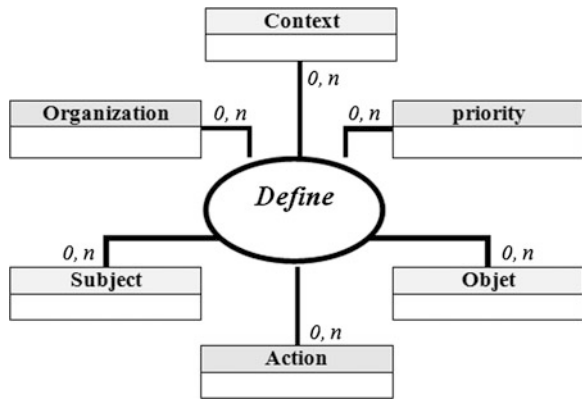


Fig. 4 Define relationship



views, the entity *Activity* corresponds to actions that partake of the same principles (e.g., *writing, consulting*).

The relationship *Consider* will be used to join up the entities *Organization*, *Action*, and *Activity*. More precisely, *Consider* (*org*, α , *a*, *p*) means that the organization *org* considers that an action α falls within the activity *a* with an uncertainty degree *p* (Fig. 3).

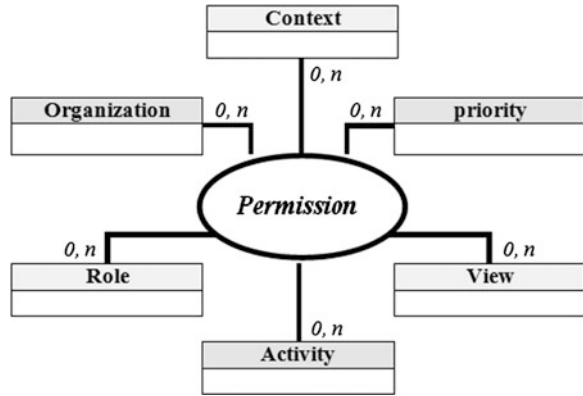
Here, *p* can reflect the uncertainty on the source that provides the permission rule. The source that is based on an old law (or national law) should have a lower certainty degree that is the one based on some recent law (or an European law).

4.5 Contexts

Contexts are used to specify the concrete circumstances where organizations grant roles permissions to perform activities on views.

The relationship *Define* (*org*, *s*, α , *o*, *c*, *p*) means that within the organization *org*, context *c* is true between subject *s*, object *o*, and action α with a certainty degree *p* (Fig. 4).

Fig. 5 Permission relationship



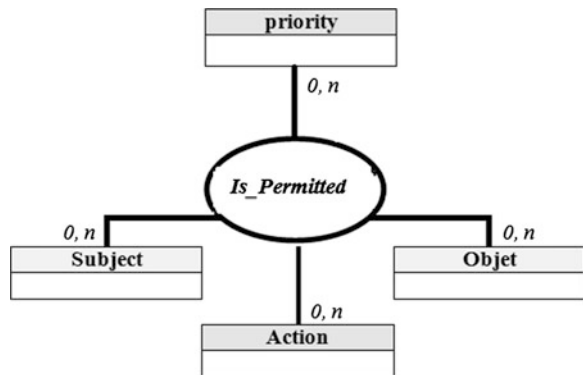
5 Prioritized Abstract and Concrete Permissions

In the OrBAC model, the relationships *Permission*, *Prohibition*, and *Obligation* correspond to relations between organizations, roles, views, activities, and contexts. The relationship *Permission* (org, r, a, v, c, p) means that the organization org grants role r permission to perform activity a on view v within the context c . The relationships *Prohibition* (org, r, a, v, c, p) and *Obligation* (org, r, a, v, c, p) are defined similarly (Fig. 5).

The relationships *Permission*, *Prohibition*, and *Obligation* are introduced as facts. They are not directly associated with users, actions, and objects but to their abstractions, respectively, roles, activities, and views.

In OrBAC, the relationship *Is_permitted* (s, α, o, p) means that that subject s is permitted to perform action α on object o with a certainty degree p . Relationships as *Is_prohibited* and *Is_obliged* are defined in the same way. These relationships are logically derived from permissions (respectively, *Prohibitions* and *Obligations*) granted to role in the following way (Fig. 6):

Fig. 6 Is_permitted relationship



If the organization *org*, within the context *c*, grants role *r* permission to perform activity *a* on view *v* with uncertainty p_1 , and if *org* employs subject *s* in role *r* with uncertainty p_2 , and if *org* uses object *o* in the view *v*, and if *org* considers that action α falls within the activity *a* and if, within the organization *org*, the context *c* is true between *s*, α , and *o*, then *s* is permitted to perform α on *o* with uncertainty $f(p_1, p_2, p_3, p_4, p_5)$.

Similarly for the *prohibition*'s relationship.

The question now is how to define $f(p_1, p_2, p_3, p_4, p_5)$? Different strategies may be defined.

5.1 Pessimistic Combination Mode

The idea in the pessimistic combination mode is to define the uncertainty degree associated with concrete permissions as the least certain degree of different elements that allow to derive such concrete permissions, namely

$$f(p_1, p_2, p_3, p_4, p_5) = \min(p_1, p_2, p_3, p_4, p_5) \quad (4)$$

This is a caution but a safe way to grant concrete permissions. One of the advantages of this combination mode is that one can directly reuse possibilistic logic machinery introduced in Sect. 2. The idea is very simple. First, entities (subjects, objects, views, roles,) are represented by unary first-order predicates, while relations (employ, use, permission, etc.) are represented by n-ary first-order predicates, depending on the number of arguments of the relations. Then, we build a possibilistic knowledge where first unary predicated get degree 1 (there is no uncertainty), while n-ary predicates of the form $Q(x_1, \dots, x_n, p_i)$ are transformed to $Q(x_1, \dots, x_n, p_i)$ where p_i is the uncertainty degree.

Once a possibilistic knowledge base is built, possibilistic logic inference [8] can be used to derive concrete permissions in an efficient way. Indeed, the computational complexity of possibilistic inference is slightly higher than the one of propositional logic. In particular, if knowledge bases only contain horn clauses, then the inference process can be achieved in a polynomial time.

5.2 Optimistic Combination Mode

The idea is to define the certainty degree associated with concrete permissions as the maximum of certainty degrees of different elements that allow to derive such concrete permissions, namely

$$f(p_1, p_2, p_3, p_4, p_5) = \max(p_1, p_2, p_3, p_4, p_5). \quad (5)$$

Clearly, this combination mode is adventurous. For instance, a user may play some rule with full certainty, and then any derived concrete permission will be fully granted, even if abstract permissions are not certain at all.

5.3 Discounting Combination Mode

The idea is to view each certainty degree as a discounting factor. The discounting effect may be obtained using the product operator, namely

$$f(p_1, p_2, p_3, p_4, p_5) = p_1 \times p_2 \times p_3 \times p_4 \times p_5 \quad (6)$$

Clearly, this combination is very caution. Besides, the transformation to possibilistic knowledge bases is not immediate and is costly.

To conclude, among the three combination modes, the pessimistic one (based on minimum operation) is the most appropriate. It provides a safe way to derive concrete permission. Moreover, it can be to easily encode in possibilistic logic from which efficient inferences can be applied.

6 Conclusion

This paper proposed an extension of OrBAC by integrating the concepts of priorities that may be associated with security policy rules. These priorities are represented in a possibility theory framework. We proposed different aggregation rules to derive the priority associated with concrete permissions from prioritized abstract permission rules. This work is important to deal with uncertain security rules, and it may also be used to solve conflicts induced by presence of permission and prohibition rules. Handling conflicts and extending our framework to deal with imprecise security rules are left for future works.

References

1. Harrison MA, Ruzzo WL, Ullman JD (1976) Protection in operating systems. *Commun ACM* 19(8):461–471
2. Lampson BW Protection. In: *Proceedings of fifth annual Princeton conference on information sciences and systems*, Princeton University, pp 437–443 March 1971
3. Sutherland D (1986) A model of information. In: *processing of the 9th national computer security conference*. National bureau of standards and national computer security center, pp 175–183 Sept 1986
4. Ferraiolo DF, Ravi S, Serban G, Richard KD, Ramaswamy C (2001) Proposed NIST standard for role-based access control. *ACM Trans Inf Syst Secur* 4(3):224–274

5. Gavrila SI Barkley JF (1996) Formal specification for role based access control user/role and role/role relationship management. Third ACM workshop on role-based, pp 81–90, 22–23 Oct 1996
6. Ravi S, Coyne EJ, Feinstein HL, Youman CE (1996) Role-based access control models. *Computer* 29(2):38–47
7. Kalam AEL, Baida REL, Balbiani P, Benferhat S, Cuppens F, Deswarte Y, Miège A, Saurel C, Trouessin G (2003) Organization based access control. 4th IEEE international workshop on policies for distributed systems and networks (Policy'03), 4–6 June 2003
8. Dubois D, Lang J, Prade H (1994) Possibilistic logic. *Handbook of Logic in artificial intelligence and logic programming*, vol 3. Oxford University Press, Oxford, pp 439–513
9. Thomas R, Sandhu R (1997) Task-based authorization controls (TBAC): a family of models for active and enterprise-oriented authorization management. 11th IFIP working conference on database security, Lake Tahoe
10. Bell DE, LaPadula LJ (1976) Secure computer systems: unified exposition and multics interpretation. Technical Report ESD-TR-73-306. The MITRE Corporation, Technical Report, March 1976
11. Biba KJ (1975) Integrity considerations for secure computer systems. Technical Report TR-3153, The Mitre Corporation, Bedford, June 1975
12. Thomas R (1997) Team-based access control (TMAC): a primitive for applying role-based access controls in collaborative environments. In: *Proceedings of the second ACM workshop on Role-based access control*, no. RBAC '97, pp 13–19
13. Cuppens F, Miège A (2003) Modelling contexts in the Or-BAC Model. 19th annual computer security applications conference (ACSAC '03), Dec 2003
14. Cuppens F, Cuppens-Boulahia N, Coma C (2006) MotOrBAC: an administration and simulation tool of security policies. Security in network architectures (SAR) and Security of information systems (SSI), first joint conference, 6–9 June 2006

Fractional Order Control for Hydraulic Turbine System Based on Nonlinear Model

Xinjian Yuan

Abstract Though fractional calculus has been used in control theory for several years, it has been applied to rotation speed control of hydraulic turbine little. In order to improve transition of hydraulic turbine under load disturbance, application of fractional order PID (FOPID) controller to hydraulic turbine governor based on nonlinear model is presented and studied. And optimal parameters are found with particle swarm optimization (PSO) algorithm. Comparisons are made with a PID governor. Simulation results show that fractional order PID controller provided better dynamics than classical PID controller, and fractional order control theory is an effective method for hydraulic turbine governor.

Keywords Hydraulic turbine · Fractional order · Particle swarm optimization · Governor · Simulation

1 Introduction

Hydraulic turbine governor is important for hydroelectric plants, and its properties are directly related to the stability of the electric power system. The main task of the turbine governor is to meet the needs of the safe operation of power generation through constantly adjusting the active power output of hydroelectric generating set. So far, mostly there are linear model-based studies on hydraulic turbine governor, and the nonlinear model-based studies are rarely reported.

Fractional calculus is an ancient branch of mathematics, the calculus of the order is extended to the scores even in complex field, fractional calculus model can improve the characterization of the dynamic system, and the application of it in the

X. Yuan (✉)

Department of Energy and Electrical, Hohai University, Nanjing, China
e-mail: Yuanxinjian11@126.com

control field is becoming a hot spot. Professor Podlubny [1] proposed $PI^\lambda D^\mu$ controller, which has a structure similar to the traditional PID controller. Cao [2] and Wu [3] apply $PI^\lambda D^\mu$ controller to pneumatic position servo control and intelligent vehicle control system, respectively, and get the desired effect. Majid [4] introduced fractional order controller to terminal voltage control and improved the robustness of the system. Wang [5] proposed $P(ID)^\mu$ controller and used in lead-lag correction design. Li [6] applied PD^μ controller for a class of second-order object control, which is simple and practical and improves the robustness of the object system. This article aims to improve the quality of turbine speed control by applying FOPID controller to the nonlinear model of hydraulic turbine.

$PI^\lambda D^\mu$ controller has five parameters: proportional gain, derivative gain, integral gain, integral order, and differential order, which improve the control strategy flexibility while tuning the controller parameters difficult, and how to select the five reasonable controls is the key to achieve excellent control. The most common methods are the dominant pole search method [7], genetic algorithm tuning [8, 9], PSO tuning [10–12], and differential evolution algorithm tuning. In this paper, particle swarm optimization (PSO) is chosen.

PSO is a global optimization algorithm, which has great probability to find the global optimal solution and has fast algorithm convergence. It has been applied in many areas and issues. In this paper, $PI^\lambda D^\mu$ controller and traditional PID controller parameters are optimized by PSO; the results show that the FOPID control system has better dynamic performance.

The specific circumstance of the nonlinear model of turbine regulating system used in this paper is shown in Sect. 2. Basic concept of fractional calculus and approximation of the differential operator is introduced in Sect. 3. PSO algorithm and method to apply PSO to design the hydraulic turbine governor are shown in Sect. 4. Experimental simulation and conclusions are shown in Sect. 5 and 6, respectively.

2 Nonlinear Model of Hydraulic Turbine

IEEE proposes a nonlinear model based on nonelastic water hammer in 1992, and the model is shown in Fig. 1.

where

- G Wicket gate;
- q Turbine flow rate;
- q_{nl} No load flow rate;
- h Turbine head;
- h_1 Head loss;
- P_m Turbine power;
- Δw Turbine speed;
- D Mechanical damping coefficient;

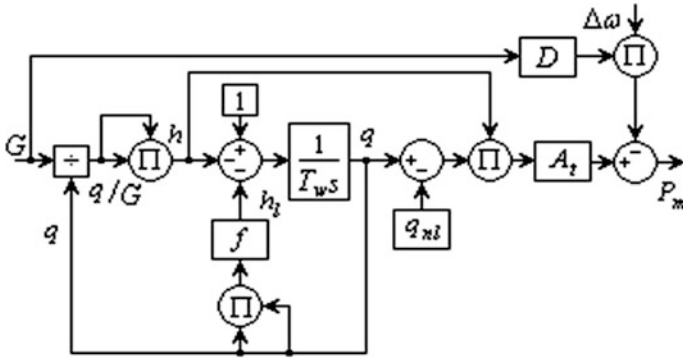


Fig. 1 Nonlinear model of hydraulic turbine system

T_w Water inertia time constant;
 f Head loss coefficient; A_t A proportionality factor.

Differential equations are as follows:

$$\begin{cases} \frac{dq}{dt} = \frac{1}{T_w} (1 - fq^2 - h) \\ h = \frac{q^2}{G^2} \\ P_m = A_t h (q - q_{nl}) - DG\Delta\omega \end{cases} \quad (1)$$

In the case of small load disturbance, approximate that

$$P_m = m_t \quad (2)$$

where m_t is turbine torque relative deviation

First-order generator model

$$\dot{x} = \frac{m_t - m_{g0} - e_n x}{T_a} \quad (3)$$

where

- m_{g0} Moment of resistance of the generator load
- e_n Controlled system self-regulation coefficient
- T_a Generator unit mechanical time
- x Turbine speed relative deviation

3 The Definition of Fractional Calculus and the Approximation of Differential Operator

3.1 Approximation of Fractional Differential Operator

Time domain definition of fractional calculus has commonly used Grünwald–Letnikov definition as well as Riemann–Liouville definition and Caputo integration.

Although the definition can be simple numerical calculation, this algorithm cannot be physical implementation. Scholars have put forward some approximate methods of fractional calculus. The most common methods are continued fractions [13], the minimum variance method [14], and famous Oustaloup approximation. Xue [15, 16] proposed a modified Oustaloup filter, which is more accurate than others, and implemented it with Simulink in Matlab. The formula of modified Oustaloup approximation in band of $[w_b, w_h]$ is shown in Eq. 4

$$s^\alpha \approx \left(\frac{dw_h}{b}\right)^\alpha \left(\frac{ds^2 + bw_h s}{d(1-\alpha)s^2 + bw_h s + d\alpha}\right) \prod_{k=-N}^N \frac{s + w'_k}{s + w_k} \tag{4}$$

where

- S Differential operator;
- α Differential order;
- b, d Positive coefficient;
- N Approximate order;
- w'_k, w_k Real poles and zeros that can be obtained through Eqs. 5 and 6.

$$w'_k = \left(\frac{dw_b}{b}\right)^{\frac{\alpha-2k}{2N+1}} \tag{5}$$

$$w_k = \left(\frac{bw_h}{d}\right)^{\frac{\alpha+2k}{2N+1}} \tag{6}$$

The approximate effect is very good in entire frequency range selected using this method.

3.2 The $PI^\lambda D^\mu$ Controller

Professor Podlubny proposed $PI^\lambda D^\mu$ controller. It is the generalization of integer order controller. Its relationship with traditional PID controller is shown in Fig. 2.

Controller introduces two factors, increased control flexibility and integer order PID controller, which are indicated by only several points in Fig. 2, and the $PI^\lambda D^\mu$

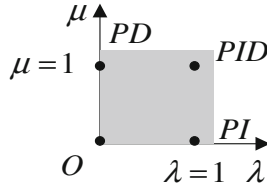


Fig. 2 The relationship between $PI^\lambda D^\mu$ controller and traditional PID controller

controller extends to the gray area and the wider first quadrant of region in Fig. 2. Actually, PID control is equivalent to special circumstances where $\lambda = 1, \mu = 1$.

4 Design of Nonlinear Fractional Order Governor with PSO Method

4.1 Construction of Speed Governor

The construction of fractional order governor is shown in Fig. 3, where

- K_p Proportional gain
- K_d Derivative gain
- K_i Integral gain
- T_n Derivative filter time constant
- T_y Wicket gate servomotor response time
- c Turbine speed relative deviation set point
- y Wicket gate relative deviation, where its relationship with wicket gate is shown in Eq. 7.

$$y = 1 - G \tag{7}$$

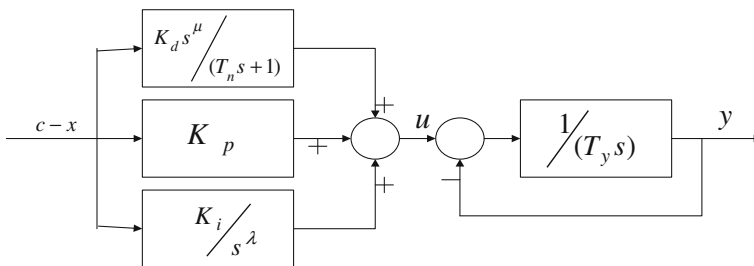


Fig. 3 Speed governor with FOPID law

The expression of control strategy is shown in Eq. 8.

$$u = (K_d + \frac{K_i}{s^\lambda} + \frac{K_d s^\mu}{T_n s + 1})(c - x) \quad (8)$$

4.2 Particle Swarm Optimization

PSO is a swarm intelligence optimization algorithm. The particle randomly generated initial velocity and position, to update their own speed and position to pursue optimal solution individual. Updated formula of standard particle swarm algorithm is as follows:

$$V_{ik}(g + 1) = w \cdot V_{ik}(g) + c_1 \cdot r_1 \cdot (P_{ik}(g) - X_{ik}(g)) + c_2 \cdot r_2 \cdot (P_{gk}(g) - X_{ik}(g)) \quad (9)$$

$$X_{ik}(g + 1) = X_{ik}(g) + V_{ik}(g + 1) \quad (10)$$

where

$k = 1, 2, \dots, d$;	$V_{ik}(g)$	The velocity of the k -th dimension of the i -th particle of the swarm;
	$P_{ik}(g)$	The individual extreme of the k -th dimension of the i -th particle of the swarm;
	$P_{gk}(g)$	The global extreme of the k -th dimension of the g -th iteration;
	$X_{ik}(g)$	The fitness of the k -th dimension of the i -th particle of the swarm;

w is linear decreased weight; c_1, c_2 non-negative constant; r_1, r_2 random number between 0 and 1.

In this paper, five parameters of $PI^\lambda D^\mu$ controller compose a particle $X_i = [K_p, K_d, K_i, \lambda, \mu]$, optimal parameter configuration through iterative search for the optimal particle. Objective function selects ITAE criterion, which is expressed as follows:

$$J = \int_0^{t_s} t \cdot |e(t)| \cdot dt \quad (11)$$

where t_s is transition time and $|e(t)|$ is speed response error absolute value; expression of w is shown in Eq. 12.

$$w(g) = w_{\max} - \frac{w_{\max} - w_{\min}}{\text{itermax}} \cdot \text{iter} \quad (12)$$

where w_{\max} is the initial inertia weight; w_{\min} the final inertia weight; iter_{\max} maximum iteration algebra; iter current iteration.

5 Simulation

5.1 Basic Simulation Parameters

Turbine system parameters of this article are from a hydropower station in China, and specific data are as follows:

Rated power $P_r = 11.0$ MW; rated head $H_r = 46.0$ m; rated flow $Q_r = 27.09$ m³/s; rated speed $n_r = 214.3$ r/min; $T_y = 0.3$ s; $T_a = 5.72$ s; $T_w = 0.83$ s; $e_g = 0$; $A_t = 1.06$; $q_{nl} = 0.95$; the parameters of $PI^\lambda D^\mu$ controller used in the simulation as well as integer order PID controller optimal with PSO. $\text{popsize} = 20$; $c_1 = 1.429$; $c_2 = 1.429$; maximum of iterations $\text{itermax} = 50$; the initial inertia weight $w_{\max} = 0.9$; the final inertia weight $w_{\min} = 0.4$; $t_s = 15$ s; the modified Oustaloup filter parameters is set to frequency lower limit $w_b = 0.001$; upper limit $w_h = 1000$; coefficient $b = 10$, $d = 9$; approximation order N set to 4.

5.2 Load Disturbance Simulation

Ten percent load disturbance parameters of fractional order governor and PID governor are shown in Table 1.

Figure 4 compares FOPID and PID controller fitness convergence curve and shows that fitness of PID controller had stable convergence at about 20 generations. Eventually, fitness of FOPID is smaller.

Figure 5 shows response of FOPID control and PID control system under load disturbance. FOPID has a smaller overshoot ratio, and the last to enter the steady-state time is also shorter, then the curve rises and falls faster. Figures 6 and 7 show the response of state variables of turbine system, the wicket gate change about 18 %, water pressure changes up to approximately 12 %, the maximum flow changes about 15 %, only the main torque change of more than 20 %, reached 23 %, a big change in the torque put forward higher requirements of the

Table 1 Governor parameters

Controller	K_p	K_d	K_i	λ	μ
PID	4.5553	2.3437	1.7584	–	–
FOPID	6.4968	3.1186	2.2719	0.9834	1.1495

Fig. 4 Fitness convergence of load disturbance

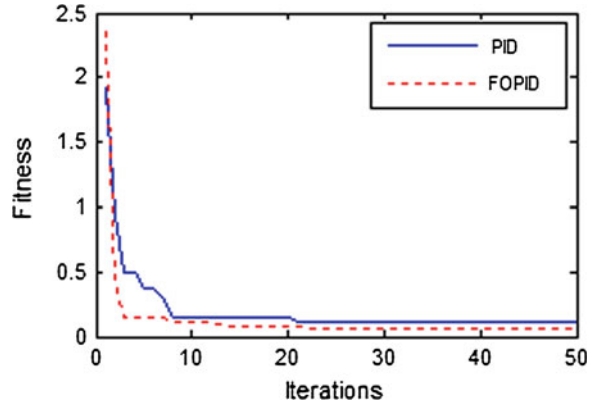


Fig. 5 Speed response of load disturbance

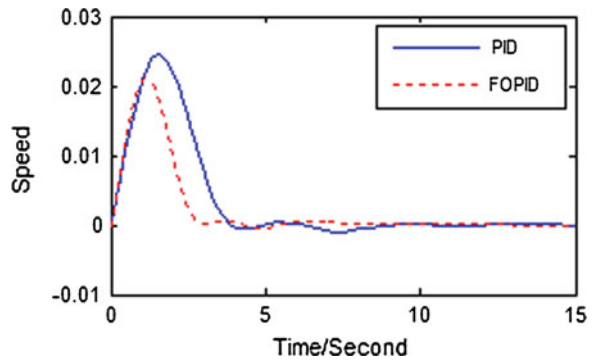
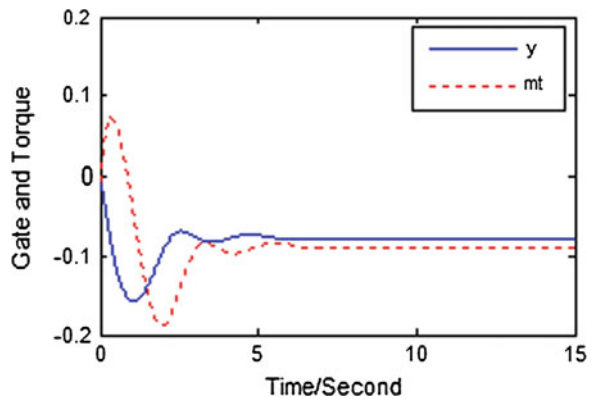
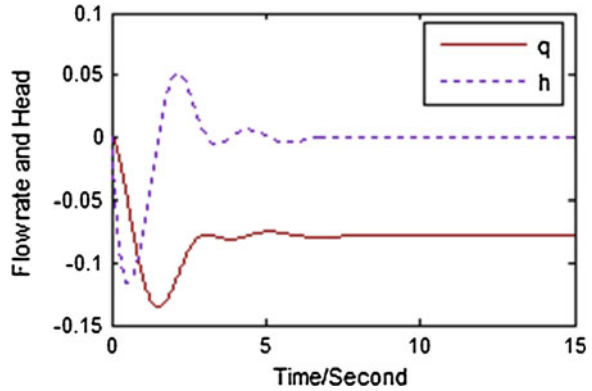


Fig. 6 Response of wicket gate and main torque



implementing agencies; on the whole, the changes in the state variables are within the acceptable range, and FOPID turbine speed regulation is effective.

Fig. 7 Response of flow rate and head



6 Summary

This paper discusses the method of how to apply fractional controllers to the nonlinear model of the turbine governor system. PSO algorithm is used to find the optimal parameters of the governor based on ITAE. Comparative simulation with integer order PID control is done, and the results show that fractional order controller is effective for turbine regulating system.

References

1. Podlubny I (1999) Fractional order systems and $PI^{\lambda}D^{\mu}$ controllers. *Trans Autom Control* 44:208–214
2. Cao JY, Cao BG, Du YT (2006) Study on application of the fractional controller to the pneumatic position servo control. *Process Autom Inst* 33:61–64
3. Wu ZY, Zhao L, Lin F (2011) Intelligent vehicle control based on fractional PID controller. *Control Eng* 18:401–404
4. Majid Z, Masoud K, Nasser S et al (2009) Design of a fractional order PID controller for an AVR using particle swarm optimization. *Control Eng Practice* pp 1380–1387
5. Wang JF, Li YK Fractional $P(ID)^{\mu}$ controller and design of lead-lag correction, *Circ Syst* 11:21–25
6. Hong S, Li YL, Chen YQ (2010) A Fractional Order Proportional And Derivative (FOPD) motion controller: tuning rule and experiments. *Trans Control Syst Technol* 18:516–520
7. Yan H, Yu SL, Li YL (2007) Fractional controller parameter design method: pole order search method. *Inf Control* 36:45–50
8. Wang C, Wang SC (2005) Search on parameters base on genetic algorithm tuning and simulation. *Comput Simul* 22:112–114
9. Biswas A, Das S, Abraham A (2009) Design of fractional-order $PI^{\lambda}D^{\mu}$ controllers with an improved differential evolution. *Artif Intell* 22:340–343
10. Cao JY, Cao BG (2006) Design of fraction order controllers based on particle swarm optimization. *Int J Control Autom Syst* 4:775–781
11. Kennedy J, Eberhart RC (1995) Particle swarm optimization. In: *Proceedings of the IEEE international conference on neural networks*. Perth, p 1942–1948

12. Shi YH, Eberhart RC (1998) A modified particle swarm optimizer. In: Proceedings of IEEE international conference on evolutionary computation, pp 69–73
13. Vinagre BM, Chen YQ et al (2003) Two direct Tustin discretization methods for fractional-order differentiator/integrator. *J Franklin Inst* 340:349–362
14. Podlubny I (2002) Geometrical and physical interpretation of fractional integration and fractional differentiation. *Fractional Calc Appl Anal* 5:357–366
15. Xue DY, Zhao CN, Chen YQ (2006) A modified approximation method of fractional order system. *IEEE ICMA*, pp 1043–1048
16. Xue DY, Zhao CN, Pan F (2006) The simulation methods and applications of fractional nonlinear systems based on the block diagram. *J Syst Simul* 18:2405–2408

Fingerprint Orientation Estimation Based on Tri-Line Model

Xiaolong Zheng, Canping Zhu and Jicheng Meng

Abstract As an intrinsic property of fingerprint, orientation field plays many important roles in fingerprint verification, including fingerprint image segmentation, fingerprint enhancement, and fingerprint matching. The image quality and singularity of fingerprint usually cause estimating error of orientation field. In this paper, an efficient orientation estimation method is proposed to address these problems, and we first coarsely estimate orientation field to give a rough description of fingerprint orientation, and then, we proposed a tri-line model in Fourier domain; according to the characteristic of fingerprint image patch, we accurately estimate fingerprint orientation field. The proposed approach has been tested on many public fingerprint databases (e.g., FVC2002), and the experimental results show that our method can achieve much more accuracy orientation field of fingerprints.

Keywords Fingerprint orientation • Tri-Line model • Fourier analysis.

X. Zheng (✉)

School of Electronics and Information, Hangzhou Dianzi University, Hangzhou, China
e-mail: xlzheng@hdu.edu.cn

C. Zhu

School of Electronic Information and Electrical Engineering,
Shanghai Jiao Tong University, Shanghai, China
e-mail: zhucanping@sjtu.edu.cn

J. Meng

School of Automation Engineering, University of Electronic Science
and Technology of China, Chengdu, China
e-mail: jcmeng@uestc.edu.cn

1 Introduction

Fingerprint verification system is the most popularly applied biometric-based security system due to the invariability and uniqueness of fingerprint feature. Generally, a fingerprint has two kinds of intrinsic feature: the local feature also called minutia provides the base of exactly fingerprint matching and the global feature which gives a rough profile of a fingerprint. As a most significant global feature, fingerprint orientation field provides rich information in fingerprint recognition, which has been used in fingerprint enhancement [1] and fingerprint matching [2] etc. Fingerprint can be viewed as the pattern of ridges and valleys on the human fingertips [3]; the trend of those ridges and valleys forms an orientation field. In the past three decades, a large number of orientation field estimation methods have been proposed [1, 4–8]. Roughly, these methods can be classified into two categories: gradient-based method and model-based method.

In nonsingular region of fingerprint, the ridges and valleys are spreading smoothly; gradient-based method takes advantage of this characteristic and estimates fingerprint orientation by using pixel gradient in fingerprint image. To get more accurate orientation field, gradient-based methods are usually followed by some post-processing, such as low-pass mean filter [1]. Gradient-based method is fast and easily implementation; however, this method is sensitive to image noise. As to fingerprint with high quality, it could achieve pretty good orientation field, while to poor images, the orientation field would be degenerated rapidly. In the singular region of fingerprint image, the flow of ridges and valleys is changed dramatically and the average gradient orientation of fingerprint patch is unsuitable in such regions.

Model-based methods estimate the fingerprint orientation with the aid of prior knowledge which is expressed by a parametric model. Ref. [4, 5] both use a kind of zero-pole model to estimate orientation field. A coarse orientation field is first obtained by using gradient-based method, and then, the model parameters are estimated from the coarse orientation field, and finally, the orientation field is generated through the model. Similar strategy is adopted in [6] where nonlinear phase portrait is used to model the orientation field. Model-based methods usually achieve good orientation field; nevertheless, they usually need some crucial information in advance, such as singular-point positions [8]. However, the reliable detection of singular point is yet still a challenge. Actually, this is a chicken-and-egg problem. For addressing this issue, Wang [9] proposed a complicated model-based method called FOMFE (fingerprint orientation model based on 2D Fourier expansions) which needs no information of singular points; on the contrast, they detect singular points using FOMFE. Recently, a slight improved version of FOMFE is proposed in [10] where two drawbacks of FOMFE have been overcome by introducing into Harris-corner strength which is helpful in removing the abrupt changes in orientation field.

Our approach is also a combination of gradient-based and model-based but avoids chicken-and-egg problem. Firstly, we calculate the coarse orientation field by gradient-based method, and then, the characteristics of fingerprint patches are analyzed in Fourier domain. We proposed a tri-line model to accurately describe the orientation of fingerprint image patches. The singular-points detection processing is avoided. Our approach is simple yet efficient in fingerprint orientation estimation.

The rest of this paper is organized as follows: The coarse orientation estimation algorithm is detailed in Sect. 2. As a core of our method, the fingerprint patch characteristics analysis and fingerprint orientation estimation method are proposed in Sect. 3. Experimental results are presented in Sect. 4. Finally, the whole paper is concluded in Sect. 5.

2 Coarse Orientation Estimation

Fingerprint is a typical flow-like image with its flows expressed by ridges and valleys, while ridges and valleys usually could only be observed with certain scales. Therefore, the orientation field of a whole fingerprint is usually calculated through the orientation of local patches. Rao [11] proposed a popular algorithm to extract orientation field from flow-like image. Fingerprint is firstly partitioned into nonoverlapping patches, and then, the domain orientation of these patches is assigned to each pixel in patches.

Supposing that $I(x, y)$ is a gray-scale fingerprint image, where x and y are the pixel coordinates. The gradient of n th pixel denoted by (I_x^n, I_y^n) can be calculated using simple *Sobel* operator or other more complex operators (e.g. *Marr-Hildreth* operator); thus, the gradient direction of n th pixel is $\theta^n = \arctan(I_y^n/I_x^n)$.

In Fig. 1, the gradient vectors are demonstrated by arrows, Fig. 1c shows that the gradients of pixels between a ridge and a valley are larger than the ones within the ridge or the valley. The normalized gradient vector on each pixel is shown in Fig. 1d. Due to the noise and gray-scale quantification, the gradient direction on each pixel in a patch is not identical. In order to describe the patch orientation as a whole, we set the domain orientation of each patch as the average gradient directions of its relative pixels to reduce the noise influence.

As shown in Fig. 1, the orientation of each pixel in one patch is diverse, we should represent the orientation of all pixels by a domain orientation. Assume the domain orientation is $(\cos \theta, \sin(\theta))$, a criterion to find such domain orientation is as follows:

$$E = \sum_n \left| (I_x^n, I_y^n) * (\cos(\theta), \sin(\theta)) \right|^2 \quad (1)$$

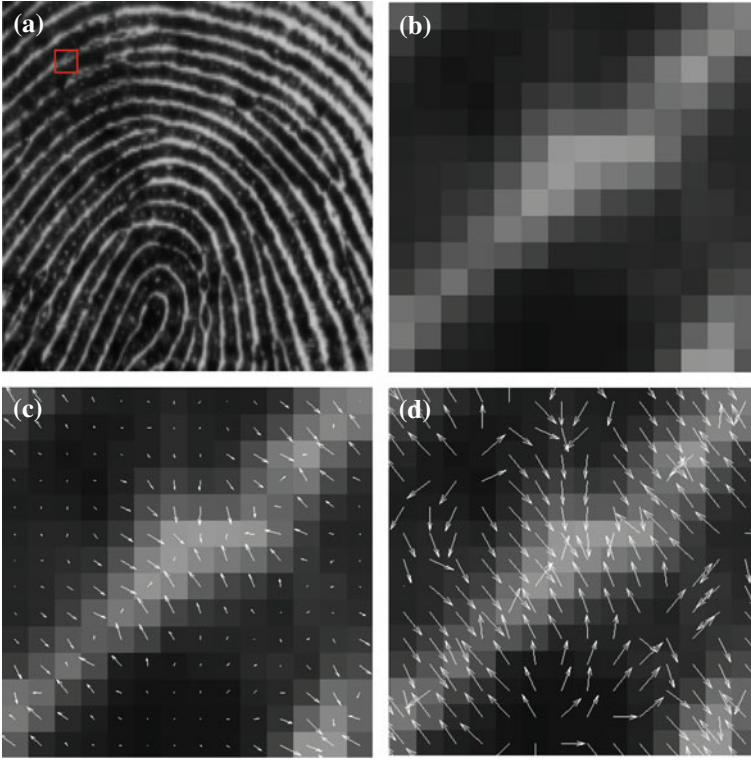


Fig. 1 The gradient vector of image patch, **a** is the original fingerprint image, **b** is the patch marked by a red rectangle in **a**, **c** visualizes the gradient vector, and **d** visualizes the direction of pixel gradient.

where (I_x^n, I_y^n) is the gradient of n th pixel in the patch whose size is $w \times w$, and the operator “ \cdot ” is the inner product between two vectors. (1) is the total energy when gradients project to the domain orientation. By maximizing (1) as to θ , the ideal domain orientation can be estimated as follows:

$$\frac{\partial E}{\partial \theta} = \sum_n (I_x^n \cos(\theta) + I_y^n \sin(\theta))(-I_x^n \sin(\theta) + I_y^n \cos(\theta)) \tag{2}$$

Let (2) be 0, we get the domain orientation as below:

$$\tan(2\theta^*) = \frac{\sum_n 2I_x^n I_y^n}{\sum_n ((I_x^n)^2 - (I_y^n)^2)} \tag{3}$$

The coherence which measures the goodness of the estimated domain orientation can be represented by the normalized energy:

$$k = \frac{\sum_n \left| (I_x^n \cos(\theta^*) + I_y^n \sin(\theta^*)) \right|}{\sum_n \sqrt{(I_x^n)^2 + (I_y^n)^2}} \tag{4}$$

From (4), $k = 1$ means the orientation of pixels is exactly identical, and therefore, the domain orientation is same as one pixel's. Theoretically, the minimum of k is 0, which means the orientation of pixels is uniformly diffused over a patch.

As to a fingerprint I , we can estimate the coarse orientation field using following algorithm:

1. Divide the original fingerprint image I into nonoverlapping patches with size as $w \times w$, in our all experiments, $w = 8$.
2. Calculate the gradient of each pixel in patch and estimate the domain orientation as follows:

$$\theta^* = \frac{1}{2} \arctan \frac{\sum_u \sum_v 2I_x(u, v)I_y(u, v)}{\sum_u \sum_v (I_x^2(u, v) - I_y^2(u, v))} \tag{5}$$

The pixel gradient depends on image quality; for a fingerprint with high quality, the gradient directions are almost identical. As a result, the domain orientation represents the orientation of patch with high accuracy. A fingerprint with high

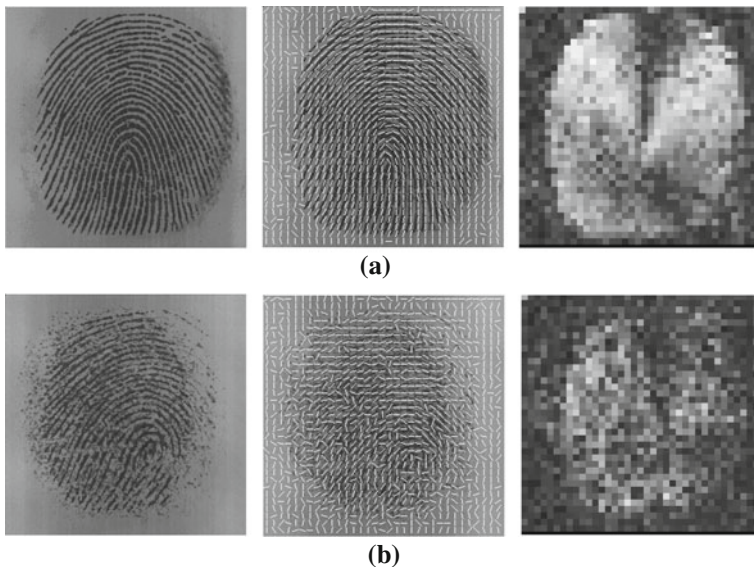


Fig. 2 a High-quality fingerprint. b Poor-quality fingerprint

image quality, shown in Fig. 2a, and a fingerprint with low image quality, shown in Fig. 2b, are used to indicate the influence of image quality on the orientation field estimation. In Fig. 2, the left two images are the original images, the middle two images indicate the orientation on each pixel, the right two images show the coherence to indicate the estimation accuracy, and the bright patch and the dark patch indicate high accuracy and low accuracy, respectively. We can see that the estimating accuracy is usually low in background since no flow appeared in these regions and the gradients are almost arbitrary. For fingerprints with low quality, since the gradient directions in patch are various, as shown in Fig. 2b, the overall coherence in the right image is low, and the above method could not give a satisfied result.

3 Tri-Line Model of Orientation Estimation

The computational results from Sect. 2 are usually rough, especially the fingerprints with low image quality. Many methods are proposed to improve the coarse estimation, including heuristic methods and model-based methods as described in Sect. 1. In this paper, we proposed a simple yet efficient approach to get accurate orientation field.

As shown in Fig. 2, the estimated orientation field of fingerprint with high image quality is fairly accurate as indicated by coherence. In those regions with much noise, the orientation coherence usually is low; for addressing this issue, we take post-process to amend the coarse orientation and re-estimate the patch orientation.

In appearance, fingerprint is composed of many flows, which spreads over fingerprint at certain frequency; therefore, in the nonsingular region of fingerprint image, the response in Fourier domain is a pair of two salient pulses due to the cyclicity of flows. However, the two pulses are deteriorated by the unavoidable noise which diffuses the frequency response, as shown in Fig. 3. Figure 3a and c are image slices from Fig. 2a and b, respectively, whose size is 32×32 to contain more local characteristic information. The image slice in Fig. 3a has high image quality; consequently, the frequency response provides two salient pulses as shown

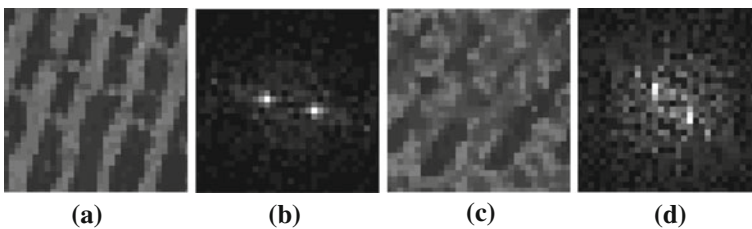


Fig. 3 Nonsingular fingerprint patch and its frequency response.

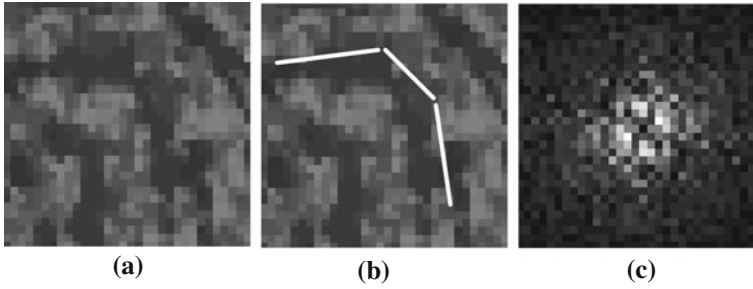


Fig. 4 Singular region and its frequency response

in Fig. 3b which are denoted by bright spots. On the other hand, the image slice in Fig. 3c is poor and Fig. 3d shows its dispersed frequency response; however, the bright spots are also salient. Comparing the spatial image with its frequency response, we have found that the line linked two brightest spots is normal to the flow orientation; therefore, we can alternatively calculate the orientation in Fourier domain.

In the normal region (nonsingular region), the above observation is reasonable. However, in the singular region, such as core or delta region in fingerprint, the flows are not parallel to each other any more. Mostly, there are three flows intersect in those regions, Fig. 4a shows an image slice from a core region; obviously, the ridges are not parallel, and there are much noise embedded into the patch, and we approach ridges with three lines as shown in Fig. 4b. Therefore, the pattern of Fig. 4a is formed by spreading three lines in parallel. Figure 4c gives the frequency response in which more than two salient pulses appear. The original image patch can be modeled by three lines; there are at least three pairs of salient pulses in Fourier domain. As to nonsingular patch, the tri-line model also holds because the other two pairs of salient pulses are too weak compared with the strongest pulse. In this paper, we model various fingerprint patches by three lines, and the domain orientation of the patch is estimated by weighting the orientation of three lines.

Based on the above tri-line model, we propose an algorithm to get accuracy fingerprint orientation field.

1. Get coarse orientation field through the algorithm described in Sect. 2, as to those patches whose coherence below a threshold th (in this paper, we assign th as 0.5), we crop a patch around the block with 32×32 .
2. Take 2D Fourier transform on the enlarged patch and obtain its frequency response.
3. Calculate the entropy of frequency response, if the entropy larger than a threshold, turn to step 4. Otherwise, re-estimate the orientation as follows:

$$\theta^* = \frac{1}{2} \arctan \left(\frac{\sum_u \sum_v w(u, v) \sin(2\theta(u, v))}{\sum_u \sum_v w(u, v) \cos(2\theta(u, v))} \right) \quad (6)$$

where $\theta(u, v)$ is the original orientation of neighbor patch at (u, v) , and $w(u, v)$ is 1 when the neighbor patch needs no re-estimation; otherwise, we let $w(u, v)$ equals to 0. Turn to step 1.

4. As to the patch with high entropy which means much salient pulse appeared in Fourier domain, we calculate the domain orientation of the patch as follows:

$$\theta^* = \frac{1}{2} \arctan \left(\frac{\sum_{n=1}^3 2w_n x_n y_n}{\sum_{n=1}^3 w_n (x_n^2 - y_n^2)} \right) \quad (7)$$

In (7), (x_n, y_n) is the position of three local maximum pulses corresponding to three lines, the position should be constrained to a half plate because of the symmetry of frequency response. Nonmaximum suppress [12] is adopted to get the position of the three local maximums. w_n is the weight of the pulse, and $w_n = f(d_n) \cdot h(g_n)$, $f(\cdot)$ is a gaussian function, $h(\cdot)$ is a linear increasing function. d_n is the distance between the local maximum pulse to the origin of Fourier domain, which measures the frequency of flows, since the flow is spreading at a certain frequency, too large distance usually corresponds to too narrow flow, on the other hand, too small distance means the interval of flows is too broad; however, these two cases are seldom appeared in practice, and thus, small weight is assigned to such two cases through the gaussian function $f(\cdot)$. g_n is the amplitude of local maximum pulse, which measures the strength of flow within three lines. For example, as to the fingerprint patch shown in Fig. 3a, there is only one pair of salient spots in Fig. 3b, the g_n of this spot is far larger than other two spots which are nearly invisible in Fig. 3b. As a result, the orientation of this patch is almost normal to the direction of the line linking two brightest spots.

4 Experimental Results

As to the two special cases in Fig. 2, Fig. 2a is a fingerprint with high image quality, while Fig. 2b has lots of noise due to the moisture of fingertip. The proposed method is applied to such two fingerprints. Experimental results are shown in Fig. 5, the resulting estimation of good fingerprint is similar as the coarse orientation field. However, for the fingerprints with low quality like the second

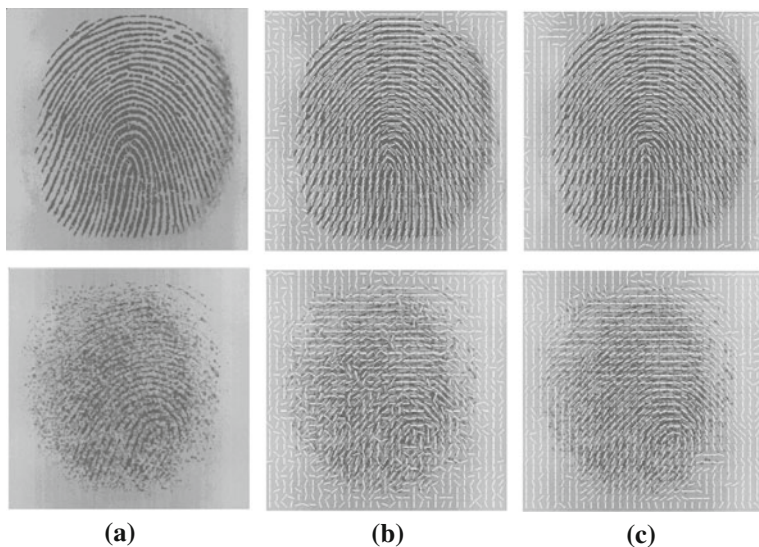


Fig. 5 The orientation fields of good and poor fingerprint image. **a** is the original image. **b** is the coarse orientation field. **c** is the orientation estimated by our method.

row in Fig. 5, our method greatly improves its estimation accuracy comparing with the coarse orientation field.

The performance of our approach has been widely tested on the public database FVC2002 DB3 [13] and a self-collection fingerprint database. FVC2002 DB3 contains 800 fingerprint images from 100 fingers (a finger provides 8 impressions) using capacitive sensor 100SC, every fingerprint image was captured at the resolution of 500dpi with the size of 300×300 . The selected testing database contains various fingerprints with different image quality, such as too dry and too wet fingerprints. The self-collection fingerprint images were gathered from 176 fingers, and each finger provides 5 impressions. The database was captured by the sensor UareU4000, and the size of fingerprint image is 356×328 . All experiments take same parameters, partly experimental results are shown in Fig. 6. Figure 6a is the original images. The first three rows are from DB3, and they are selected as too wet fingerprints. The last three rows are selected from self-collection database as too dry fingerprints, and the gray-scale mean of the last two images have been adjusted for demonstration. The original images have been cropped to shown the main foreground of fingerprint. Figure 6b is the coarse orientation field produced by the algorithm in Sect. 2. Figure 6c demonstrates the results of the proposed method. Obviously, our method gives more accurate estimation over coarse field.

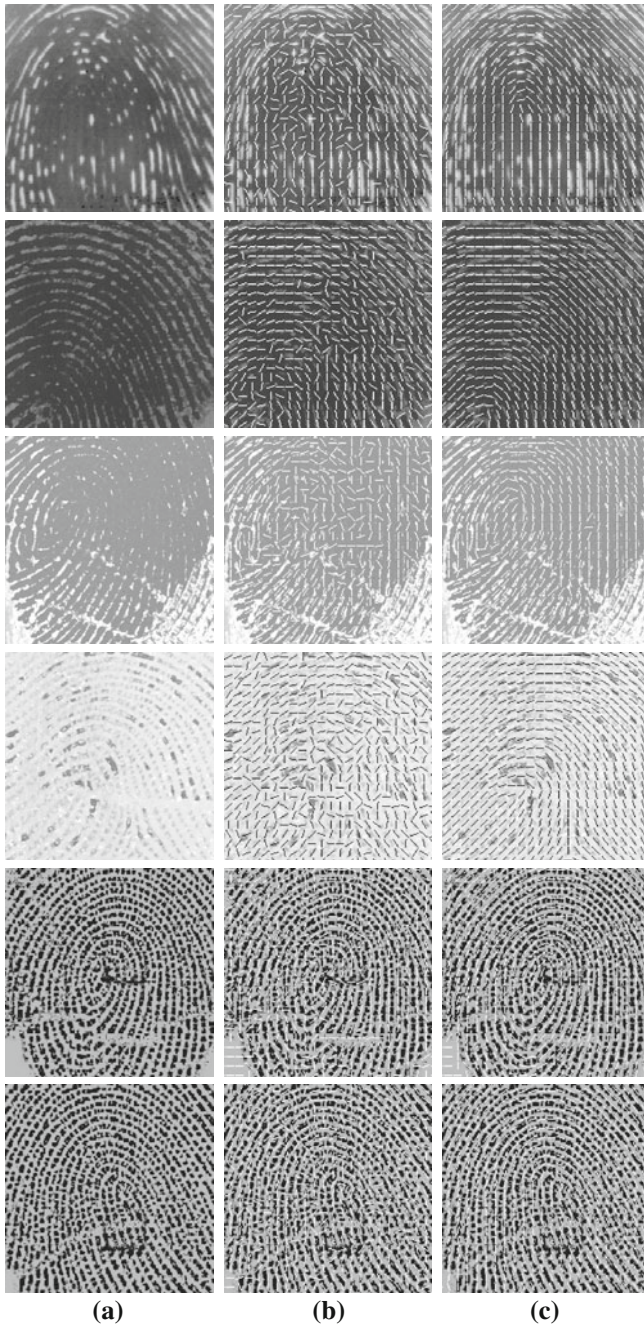


Fig. 6 The experimental examples of orientation field. **a** The original images. **b** The coarse orientation field. **c** The results of our method.

5 Conclusion

In this paper, we proposed a simple yet efficient algorithm to estimate fingerprint orientation field. Fingerprint image can be viewed as an assembly of flows. In the nonsingular region, the flows are arranged in parallel, and the frequency response gives two salient spots. However, in the singular region, the response is much more complicated. We describe these two different cases by an united model, tri-line model, with the assumption that every fingerprint patch is composed of three flow lines; therefore, in the Fourier domain, three pairs of local maximum be detected to give a compound direction which is normal to the orientation of the local patch. The tri-line model is intuitional and suitable to singular regions. As to nonsingular regions, only one flow line can be observed; nevertheless, the tri-line is also suitable for this case, since the other two flow lines are too weak to be observed. In the Fourier domain, the other two pairs of spots have slight influence on the patch orientation which mainly determined by the direction of the most salient spots. The experiments on the public fingerprint database demonstrate that the proposed method obtains the improvement on the estimating accuracy.

Acknowledgments This work is supported in part by the National Science Foundation for Distinguished Young Scholars of China under Grant No. 60905011 and the Startup Funding of Hangzhou Dianzi University(No. KYS045609059).

References

1. Hong L, Wan Y, Jain A (1998) Fingerprint image enhancement: algorithm and performance evaluation. *IEEE Trans Pattern Anal Mach Intell* 20(8):777–789
2. Tico M, Kuosmanen P (2003) Fingerprint matching using an orientation-based minutia descriptor. *IEEE Trans Pattern Anal Mach Intell* 25(8):1009–1014
3. Jain A, Prabhakar S, Hong L, Pankanti S (May 2000) Filterbank-based fingerprint matching. *IEEE Trans Image Process* 9(5):846–859
4. Gu J, Zhou J, Zhang D (2004) A combination model for orientation field of fingerprints. *Pattern Recognit* 37:543–553
5. Sherlock B, Monro D (1993) A model for interpreting fingerprint topology. *Pattern Recognit* 26:1047–1095
6. Yau W, Jun L, Han W (2004) Nonlinear phase portrait modeling of fingerprint orientation. In: *Proceedings of IEE 8th control, automation, robotics and vision conference*, pp 1262–1267
7. Li J, Wei Y (2004) Prediction of fingerprint orientation. In: *Proceedings of 17th international conference on pattern recognition*, pp. 436–439
8. Chen X, Tian J, Zhang Y, Yang X (2005) A robust orientation estimation algorithm for low quality fingerprints. In: *Proceedings of advances in biometric person authentication*, pp 95–102
9. Wang Y, Hu J, Phillips D (2007) A fingerprint orientation model based on 2D Fourier expansion (FOMFE) and its application to singular-point detection and fingerprint indexing. *IEEE Transac Pattern Anal Mach Intell* 29(4):573–585
10. Tao X, Yang X, Cao K, Wang R, Li P, Tian J (2010) Estimation of fingerprint orientation field by weighted 2D Fourier expansion model. In: *Proceedings of 20th international conference on pattern recognition*, IEEE Computer Society, Washington, pp 1253–1256

11. Rao A (1990) A taxonomy for texture description and identification. Springer, Berlin
12. Devernay F (1995) A non-maxima suppression method for edge detection with sub-pixel accuracy. Technical report 2724, INRIA Sophia-Antipolis
13. Maio D, Maltoni D, Cappelli R, Wayman J, Jain A (2002) “Fvc 2002: second fingerprint verification competition,” in Proceedings of 16th International Conference on Pattern Recognition, IEEE Computer Society, Quebec City, Canada, pp 811–814

Joint Rate and Power Allocation for Cognitive Radios Networks with Uncertain Channel Gains

Zhixin Liu, Jinfeng Wang, Hongjiu Yang and Kai Ma

Abstract In this paper, we consider a cognitive radio system with multiple cognitive radio (CR) links in the same neighborhood and propose an opportunistic power control strategy for the CR links. The key feature of the proposed strategy is that, via opportunistically adjusting the CR links' transmit power, the cognitive user can maximize its achievable transmission rate without degrading the outage probability of the primary user. Since it is difficult to track channel gains instantaneously for dynamic cognitive radio network in practice, the case where only mean channel gains averaged over short-term fading are available is considered. Under such scenarios, we derive interference constraints violation probabilities for primary user so that the interference constraints are only violated with desired limit. The achievable sum rate of the CR links under the proposed power control strategy is analyzed and simulated, taking into the impact of imperfect channel estimation. The robustness of the network system is improved with the proposed scheme.

Keywords Cognitive radios networks · Uncertain channel gains · Multiple cognitive radio links.

1 Introduction

Fixed access and resource allocation approaches to cope with interference have led to many drawbacks, such as scarcity of the available bandwidth, expensive licenses, and under-utilization of the spectrum in space, frequency, and time [1]. Moreover, with the wide-spread deployment of various wireless communication systems, radio spectrum has become a scare resource in particular. Being widely considered as a promising solution to this dilemma, cognitive radio is able to dramatically improve the spectrum utilization by allowing secondary (unlicensed)

Z. Liu (✉) · J. Wang · H. Yang · K. Ma

Institute of Electrical Engineering, Yanshan University, Qinhuangdao, Hebei, China
e-mail: lzxauto@ysu.edu.cn

users to opportunistically or concurrently access spectrum allocated to primary (licensed) users [2, 3].

In the cognitive protocols proposed earlier, only when particular frequency bands are not concurrently used by any primary users, the cognitive users can transmit in these channels [4, 5]. In this scenario, since some interference from secondary user transmissions will not be harmful to the inactive primary users, it is not necessary to impose severe restrictions on the transmission power of the secondary users. In some recent studies, secondary users are allowed to transmit simultaneously in the same frequency band with the primary users. This imposes severe restrictions on the transmission power of the secondary users, so as not to cause any harmful interference to the active primary users [6, 7]. Under the constraints that no interference was created for the primary user and the primary encoder-decoder pair was oblivious to the presence of cognitive radio, the capacity of the cognitive user was estimated in Jovicic and Viswanath [8]. The authors in Ghasemi and Sousa [9] proposed a power allocation scheme in fading environment, which maximizes the capacity of the secondary user given the interference temperature constraint at the primary receiver. The problem of optimal power control for secondary users under interference constraints for primary users was formulated as a concave minimization problem in Wang et al. [10].

There has been lots of works in the literature maximizing the sum rate of the secondary users in a dynamic spectrum access environment. While most of the proposed power control strategies are depend on the instantaneous channel state information (CSI) and assume that the cognitive user is able to gather perfect CSI in Hamdi et al. [11], a transmit power control scheme for a secondary user was proposed, which exploited the location information of the primary receiver obtained indirectly through spectrum sensing, to reduce the interference to the primary receiver. An alternative way to protect the primary user's transmission and to realize spectrum sharing between the primary user and the cognitive users is proposed in Kim et al. [12]. Considering imperfect channel estimation, the author also proposes a modified power control strategy with protection gap to reduce the sensitivity of its strategy to the estimation errors.

In this paper, we consider the case where instantaneous channel gains are not available, and only mean channel gains from secondary users to primary receiving points are available. Aimed at maximizing the sum rate of the secondary users, we propose an efficient power control strategy under the interference power constraints and individual transmit power constraints.

2 System Model and Problem Formulation

As shown in Fig. 1, we consider an OFDM-based cognitive network including one primary user and CR links, and each link can access to n orthogonal channels. Assuming σ_k^i is circular symmetric complex Gaussian signals and noise. The signal to interference and noise ratio (SINR) of user i in channel k is given by

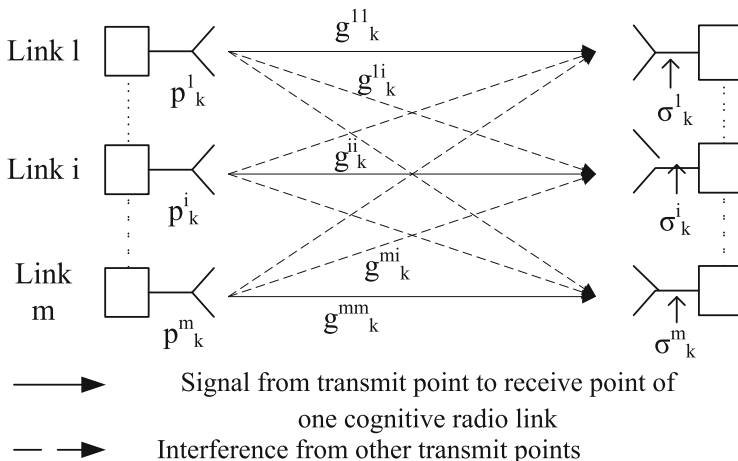


Fig. 1 System model with m CR links for a channel k

$$\gamma_k^i = \frac{g_k^{ii} p_k^i}{\sigma_k^i + \sum_{j=1, j \neq i}^m g_k^{ij} p_k^j} \tag{1}$$

where g_k^{ij} denotes the link gain from the transmitting point of cognitive radio link j to the receiving point of link i , g_k^{ii} denotes the link gain between cognitive radio link i is transmitting point and receiving point, p_k^i denotes the transmit power of link i at the channel k .

Consequently, the rate is modeled by the bandwidth normalized Shannon capacity

$$R_k^i = \log \frac{g_k^{ii} p_k^i}{\sigma_k^i + \sum_{j=1, j \neq i}^m g_k^{ij} p_k^j} \tag{2}$$

We formulate the power and rate allocation framework that aim at maximizing the sum rate. In this problem, secondary users are subjected to individual transmit power constrains as well as interference power constraints. Mathematically, the problem can be formulated as follows:

$$\begin{aligned} \max \quad & \sum_{i=1}^m \sum_{k=1}^n \log \left(1 + \frac{g_k^{ii} p_k^i}{\sigma_k^i + \sum_{j=1, j \neq i}^m g_k^{ij} p_k^j} \right) \\ \text{s. t.} \quad & \sum_{k=1}^n p_k^i \leq P_{\max}^i \quad i = 1, \dots, m \\ & \sum_{i=1}^m p_k^i h_k^i \leq I_k \quad k = 1, \dots, n \end{aligned} \tag{3}$$

where P_{\max}^i is the power threshold for secondary link i , h_k^i denote the instantaneous channel gain from the transmitting point of secondary link i to primary user at sub-channel k , and I_k is the interference threshold for the primary user on channel k . We assume that the secondary user can dynamically track the sum interference from other secondary links at its receiving point.

As mentioned above, we would like to perform joint rate and power control strategy under two constrains: individual power constrains for CR links and interference constrains for primary user. It is difficult to estimate instantaneous channel gains from secondary transmitting points to primary receiver. However, secondary transmitting points can estimate the mean channel gains from primary transmitter to themselves by exploiting pilot signals transmitted from primary transmitter. Let h_k^i denote the instantaneous channel gain from the transmitting point of secondary link i to primary user at sub-channel k , and \bar{h}_k^i be its mean value averaged over short-term fading. The interference power from secondary link received by the primary user is characterized by A , and I_k is the interference threshold for the primary user at channel k . Then, the interference constrains can be written as

$$A = \sum_{i=1}^m p_k^i h_k^i \tag{4}$$

$$\eta_k = A \leq I_k, k = 1, \dots, n \tag{5}$$

With only mean channel gains, interference constrains can be set in an average sense as follows:

$$\bar{\eta}_k = \sum_{i=1}^m p_k^i \bar{h}_k^i \leq \alpha I_k, k = 1, \dots, n \tag{6}$$

where $\alpha < 1$.

Therefore, the optimization problem in Haykin [3] can be written as [7]

$$\begin{aligned} \max \quad & \sum_{i=1}^m \sum_{k=1}^n \log\left(1 + \frac{g_k^{ii} p_k^i}{\sigma_k^i + \sum_{j=1, j \neq i}^m g_k^{ij} p_k^j}\right) \\ \text{s.t.} \quad & \sum_{k=1}^n p_k^i \leq P_{\max}^i, i = 1 \dots n \\ & \sum_{i=1}^m p_k^i \bar{h}_k^i \leq \alpha I_k, k = 1 \dots n \end{aligned} \tag{7}$$

3 Joint Rate and Power Allocation Algorithm

Since the instantaneous interference level η_k may exceed the tolerable limit I_k , and it will violate the absolute interference constraint, i.e., $\eta_k \leq I_k$. We define a constraint on the violation probability as follows:

$$\Pr[\eta_k > I_k | \bar{\eta}_k \leq \alpha I_k] \leq \delta^{(I)} \quad k = 1, \dots, n \tag{8}$$

where $\delta^{(I)}$ denotes the maximum interference violation probability allowed for primary receiving point.

We define the short-term fading $a_k^i = h_k^i / \bar{h}_k^i$ and assume it is exponentially distributed with $p.d.f. f_{a_k^i}(x) = a_k^i e^{-x/a_k^i}$. We have the following proposition,

Proposition The violation probability for the average-sense interference constraint is evaluated as

$$\Pr[\eta_k > I_k | \bar{\eta}_k \leq \alpha I_k] = \sum_{i=1}^m \pi_k^i \exp \left[-\frac{I_k}{\bar{h}_k^i p_k^i} \right] \tag{9}$$

s. t. $\bar{\eta}_k \leq \alpha I_k \quad k = 1, \dots, n$

where $\pi_k^i = \prod_{l=1, l \neq i}^m \frac{\bar{h}_k^l p_k^l}{\bar{h}_k^l p_k^l - \bar{h}_k^i p_k^i}$.

Proof The tolerable interference limit η_k can be expressed as (10)

$$\eta_k = \sum_{i=1}^m h_k^i p_k^i = \sum_{i=1}^m a_k^i \bar{h}_k^i p_k^i \tag{10}$$

It can be shown that it has the *p.d.f.*

$$f_{\eta_k}(x) = \sum_{i=1}^m \frac{\pi_k^i}{\bar{h}_k^i p_k^i} e^{-x/\bar{h}_k^i p_k^i} \tag{11}$$

The violation probability is evaluated as

$$\Pr[\eta_k > I_k | \bar{\eta}_k \leq \alpha I_k] = \int_{I_k}^{\infty} f_{\eta_k}(x) dx = \sum_{i=1}^m \pi_k^i \exp \left[-\frac{I_k}{\bar{h}_k^i p_k^i} \right] \tag{12}$$

The proof is end.

Therefore, with the opportunistic constraint, the optimization problem (6) can be written as (13). Based on the Lagrangian dual method, we have Lagrangian parameter $\lambda = [\lambda^1, \dots, \lambda^m]$ and $\mu = [\mu_1, \dots, \mu_k]$. Let

$P = [p_1^1; p_2^1; \dots; p_n^1; \dots; p_1^m; p_2^m; \dots; p_n^m]$ denotes the allocated power. Then, the Lagrangian function of the optimal problem (7) can be formulated as (14)

$$\begin{aligned}
& \max \sum_{i=1}^m \sum_{k=1}^n \log \left(1 + \frac{g_k^{ii} p_k^i}{\sigma_k^i + \sum_{j=1, j \neq i}^m g_k^{ij} p_k^j} \right) \\
& \text{s.t. } \sum_{k=1}^n p_k^i \leq P_{\max}^i \quad i = 1, \dots, n \\
& \quad \sum_{i=1}^m p_k^i \bar{h}_k^i \leq \alpha I_k \quad k = 1, \dots, n \\
& \quad \Pr[\eta_k > I_k | \bar{\eta}_k \leq \alpha I_k] \leq \delta^{(l)} \quad k = 1, \dots, n
\end{aligned} \tag{13}$$

$$\begin{aligned}
L(P, \lambda, \mu) &= \sum_{i=1}^m \sum_{k=1}^n \log \left(1 + \frac{g_k^{ii} p_k^i}{\sigma_k^i + \sum_{j=1, j \neq i}^m g_k^{ij} p_k^j} \right) \\
&+ \sum_{i=1}^m \lambda^i \left(P_{\max}^i - \sum_{k=1}^n p_k^i \right) \\
&+ \sum_{k=1}^n \mu_k \left(\alpha I_k - \sum_{i=1}^m p_k^i h_k^i \right)
\end{aligned} \tag{14}$$

The Karush-Kuhn-Tucker (KKT) conditions for the user i and $\forall k = 1, \dots, n$ are obtained by applying $\frac{\partial L}{\partial p_k^i} = 0$, which is

$$\frac{\partial L}{\partial p_k^i} = \frac{g_k^{ii}}{\sigma_k^i + \sum_{j=1}^m g_k^{ij} p_k^j} - \lambda^i - \mu_k \times h_k^i = 0 \tag{15}$$

Note that if the power quantities p_k^j with $j \neq i$ are all fixed, then the interference from other secondary transmitters $q_k^i = \sum_{j=1, j \neq i}^m g_k^{ij} p_k^j$ are constants independent of p_k^i .

Therefore, we have

$$p_k^{i*} = \frac{1}{\lambda^i + \mu_k \times h_k^i} - \left(\sigma_k^i + \sum_{j=1, j \neq i}^m g_k^{ij} p_k^j \right) \tag{16}$$

3.1 The Per Link Sum-power Constraints

Since the power allocation follows a water-filling approach over the channels, a sum-power exceeding the permissible one would require reducing each channel's power with the same amount, since they are controlled by the same Lagrangian parameter λ^i . According to the sub-gradient algorithm, the sub-gradient of λ^i is written as

$$\Psi^i(\lambda^i) = P_{\max}^i - \sum_{k=1}^n p_k^i \quad (17)$$

3.2 The Per Channel Interference Constraints

For the purpose of not cause harmful interference to primary user when violate the probability, $\delta^{(l)}$ would reduce each secondary links power similarly as described in A. The sub-gradient of μ_k is written as

$$\Phi_k(\mu_k) = \alpha I_k - \sum_{i=1}^m p_k^i h_k^i \quad (18)$$

Therefore, we can update λ^i and μ_k as follows

$$\lambda^i(t+1) = [\lambda^i(t) - \beta(t)\Psi^i(\lambda^i)]^+ \quad (19)$$

$$\mu_k(t+1) = [\mu_k(t) - \beta(t)\Phi_k(\mu_k)]^+ \quad (20)$$

where $[x]^+ = \max(0, x)$ and $\beta(t)$ satisfies

$$\lim_{t \rightarrow \infty} \beta(t) = 0, \sum_{t=1}^{\infty} \beta(t) = \infty \quad (21)$$

The convergence of the proposed algorithm is guaranteed by applying the monotonically increasing property of each step and the convexity of the problem. The flow chart of the algorithm is given in Fig. 2.

Algorithm Joint rate and power allocation with desired interference violation probabilities

1. Initialized $\alpha = 1$,
2. Solve the joint rate and power allocation problem with current values of α .
3. Calculate the violation probability for interference constraints using proposition and check whether they are smaller than the desired values in (2.7). If yes, finish; otherwise go to step 4.

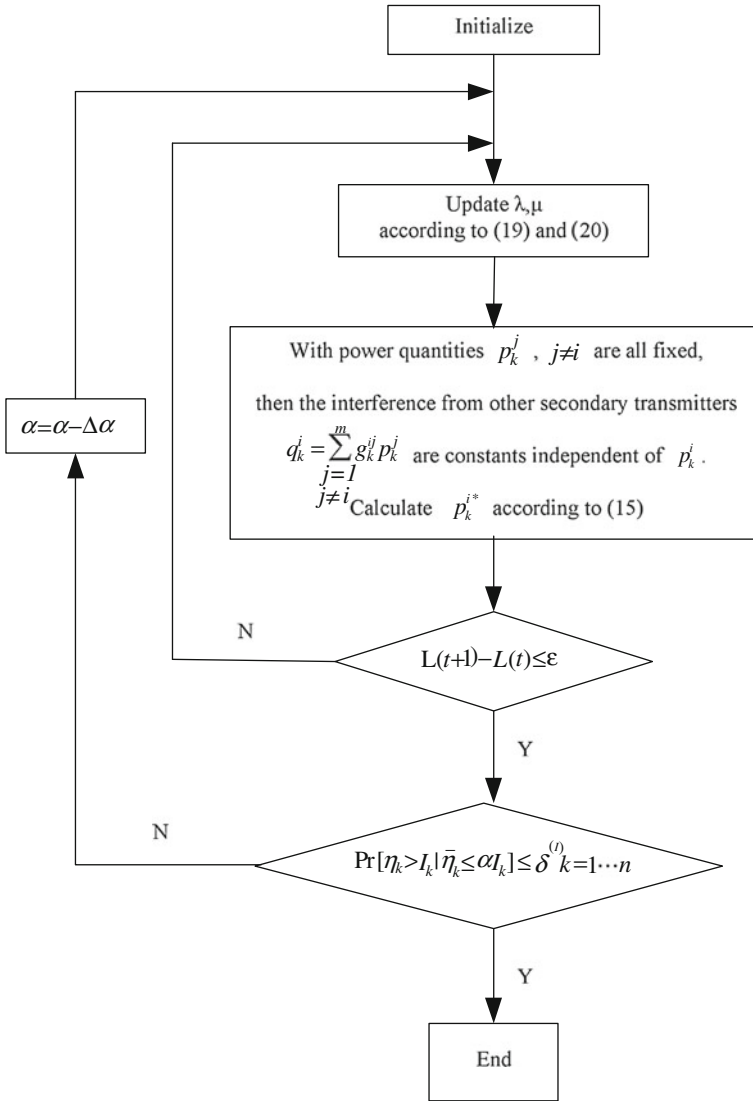


Fig. 2 Flow chart for the joint rate and power allocation algorithm

- Adjust the conservative factor as follows. If any constraints in (2.7) are violated, perform the following update

$$\alpha = \alpha - \Delta\alpha \tag{22}$$

where $\Delta\alpha$ is small adjustment value.

- Return to step 2.

4 Simulation Results

In this section, we provide numerical results to illustrate the effectiveness of the proposed algorithm. We consider a single cell where there are 5 secondary links and 3 sub-channels, a path loss exponent of 4 to 5, and Rayleigh fading. The power mask assumes for simplicity the same values for all users. All performance measures are obtained by averaging over 100 simulation runs. The adjustment value for conservative factor in Algorithm is chosen to be $\Delta\alpha = 0.05$.

As presented in our paper, we derive interference constraints violation probabilities for primary user. The performance comparison over different values of I_k in terms of the interference constraints violation probability is given.

As shown in Fig. 3, without the violation probability constrains $\Pr[\eta_k > I_k | \bar{\eta}_k \leq \alpha I_k] \leq \delta^{(l)} k = 1, \dots, n$ and the adjustment $\alpha = \alpha - \Delta\alpha$, the interference constrains violation probabilities are larger than the desired limit 0.1. In the other hand, we can make Pr 1, Pr 2, and Pr 3 under the limit using our algorithm. This successfully protects the primary signals as we mentioned before.

We show the interference constraints violation probability Pr and α in Fig. 4 for different values of $\delta^{(l)}$. As expected, the interference constraints violation probability Pr and α decrease as the maximum interference constrains violation probability $\delta^{(l)}$ becomes more stringent. In this way, we can adapt the power of the secondary links according to the maximum interference constrains violation probability $\delta^{(l)}$, so that not causing harmful interference to PU.

The performance comparison over different values of I_k in terms of the sum rate is shown in Fig. 5. We can conclude that though it is slightly lower achievable sum rate than that without violation probability constrains as expected, it can supply more sufficient protection for primary users.

We show the convergence of the proposed algorithm in Figs. 6 and 7. Here we set power constraints $P_{\max} = 0.1$ and interference constraints $I_k = 4.4 \times 10^{-6}$.

Fig. 3 P_r of the two algorithms versus interference constraints

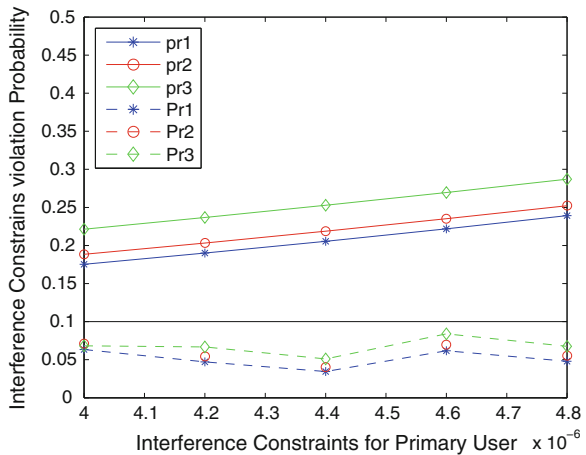


Fig. 4 $\delta^{(l)}$ versus interference constraints violation probabilities P_r and α

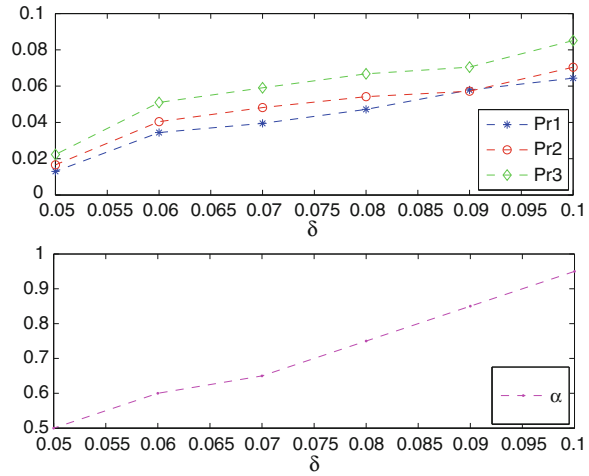


Fig. 5 Sum rate of the two algorithms versus interference constraints

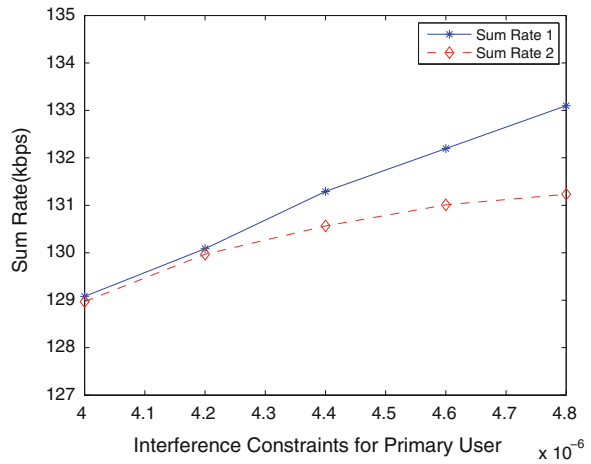


Fig. 6 λ versus number of iteration

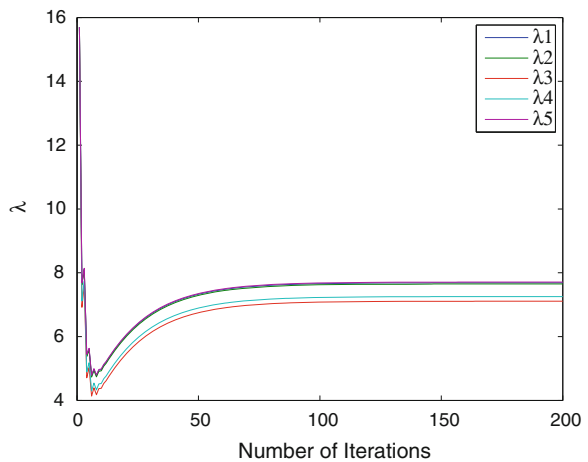
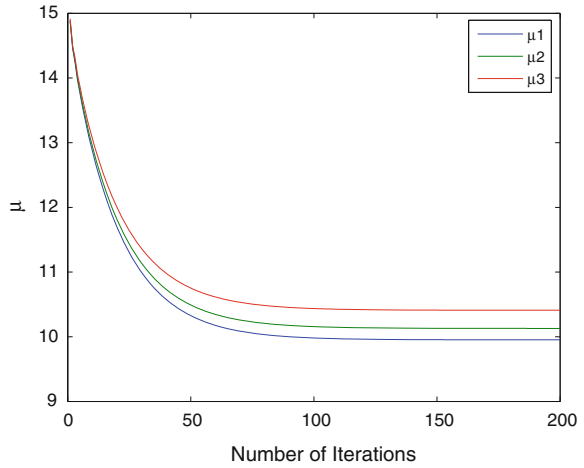


Fig. 7 μ versus number of iteration



5 Conclusions

This paper has proposed an efficient algorithm to optimally solve the sum rate maximization problem under individual transmit power constrains and interference constrains. We consider the case where only mean channel gains averaged over short-term fading are available since tracking channel gains instantaneously for dynamic cognitive radio network may be very difficult in practice. For the purpose of not causing harmful interference to PU, we derive interference constraints violation probabilities for primary user. A globally optimal solution has been obtained by solving the proposed algorithm. This work was supported partially by the NSFC under Grant 61104033 and 61174127; the NSFC key Grant with No. 60934003; and the Hebei Provincial Natural Science Fund under Grand F2012203109 and F2012203126. The work of Hongjiu Yang was supported by the National Natural Science Foundation of China under Grant 61203023 and the Postdoctoral Science Foundation of China under Grant 2012M510769, respectively.

References

1. FCC (2002) FCC spectrum policy task force report. ET-Docket 20–135, November. FCC, Washington, DCJ. In: Maxwell C (ed) A treatise on electricity and magnetism, 3rd ed, vol 2. Clarendon, Oxford, 1892, pp 68–73
2. Mitola J III, Maguire GQ (Aug 1999) Cognitive radios: making software radios more personal. IEEE Personal Commun 6(4):13–18
3. Haykin S (Feb 2005) Cognitive radio: brain-empowered wireless communications. IEEE J Select Areas Commun 23(2):201–202
4. Brodersen RW, Wolisz A, Cabric D et al (2004) Corvus: a cognitive radio approach for usage of virtual unlicensed spectrum. UC Berkeley White Paper, July

5. Horne WD Adaptive spectrum access: using the full spectrum space. In: Proceedings of the 31st annual telecommunication policy research conference. Available <http://tprc.org>
6. Srinivasa S, Jafar SA (2007) The throughput potential of cognitive radio: a theoretical perspective. *IEEE Commun Mag* 45(5):2637–2642
7. Zhang L, Liang YC, Xin Y (Jan 2008) Joint beamforming and power allocation for multiple access channels in cognitive radio networks. *IEEE J Select Areas Commun* 26(1):38–51
8. Jovicic A, Viswanath P (2006) Cognitive radio: an information-theoretic perspective. *IEEE Trans Inform Theor*
9. Ghasemi A, Sousa ES (Feb 2007) Fundamental limits of spectrum-sharing in fading environments. *IEEE Trans Wirel Commun* 6(2):649–658
10. Wang W, Peng T, Wang W (2007) Optimal power control under interference temperature constraints in cognitive radio network. In: Proceedings of the IEEE, wireless communications and networking conference (WCNC'07), pp 116–120, March 2007
11. Hamdi K, Zhang W, Letaief KB (2007) Power control in cognitive radio systems based on spectrum sensing side information. In: Proceedings of the IEEE international conference on, communications (ICC'07), pp. 5161–5165, June 2007
12. Kim DI, Le DI, Hossain E (Dec 2008) Joint rate and power allocation for cognitive radios in dynamic spectrum access environment. *IEEE Trans Wirel Commun* 7:5517–5527

Link Prediction Based on Sequential Bayesian Updating in a Terrorist Network

Cheng Jiang, Juyun Wang and Hua Yu

Abstract Link prediction techniques are being increasingly employed to detect covert networks, such as terrorist networks. The challenging problem we have been facing is to improve the performance and accuracy of link prediction methods. We develop an algorithm based on Sequential Bayesian Updating method that combines probabilistic reasoning techniques. This algorithm adopts a recursive way to estimate the statistical confidence of the results a priori and then regenerate observed graphs to make inferences. This novel idea can be efficiently adapt to small datasets in link prediction problems of various engineering applications and science researches. Our experiment with a terrorist network shows significant improvement in terms of prediction accuracy measured by mean average precision. This algorithm has also been integrated into an emergency decision support system (NBCDSS) to provide decision-makers' auxiliary information.

Keywords Link prediction · Probabilistic reasoning · Bayesian inference · Decision-making · Terrorist network

1 Introduction

Link prediction is a fundamental but important task in data mining and is applied in various areas. It is employed to look insight into the structure patterns, capabilities, and vulnerabilities of our interested organizations [1], to organize

C. Jiang · H. Yu (✉)

College of Engineering, University of Chinese Academy of Sciences, Beijing, China
e-mail: yuh@ucas.ac.cn

C. Jiang

e-mail: jiangcheng10@mails.ucas.ac.cn

J. Wang

The School of Science, Communication University of China, Beijing, China
e-mail: wangjuyun@cuc.edu.cn

unstructured data in a linked pattern in response to query quickly in information retrieval [2], to recognize identical entities with information from different data sources in recommend systems [3], and to predict the interactions between proteins in biology information networks [4, 5]. The methods of link prediction can be mainly separated into three categories [6].

Methods like information content [7], cosine coefficient [8], and mutual information [9] are based on the similarity of pairs of nodes. These methods seek to find effective similarity measures between nodes as their targets, comparing these measures to a fixed threshold in order to judge whether there exists a link between the nodes.

Methods such as common neighbors [10], Jaccard coefficient [11], Adamic–Adar [12], and preferential attachment [13] are based on topological structure, and these methods apply graph theory and social network analysis technology to make scores for pairs of nodes based on local topological structure or global topological structure.

The third category of methods aims to abstract the underlying structure from the observed data and network to a compact probabilistic model according to some optimization strategies such as maximum likelihood or maximum a posterior probability. Then, the missing links could be regenerated using the learned model. This type of methods includes Bayesian inference [14], hierarchical Bayesian inference [15], and stochastic relational model [16].

In this paper, we propose an algorithm based on sequential Bayesian updating that combines probabilistic reasoning techniques. It can make best use of limited observed data and obtain satisfied accuracy through recursive inferences. Therefore, it can work well against situations of small datasets in link prediction problems. We test this algorithm with the (believed-defunct) Greek terrorist organization “Revolutionary Organization November 17” (N17). The experiment shows that the accuracy raised by almost 10 % with our algorithm, compared to the Bayesian inference put forward by Rhodes with the same datasets.

2 Background and Related Work

Terrorist activities occur frequently in our world and bring out tremendous hazards, such as Sarin gas attack on the Tokyo subway in 1995, 9–11 attack in America, and suchlike. Counter-terrorist experts and decision-makers would benefit from the information and take effective measures against attacks, if we found link interactions within the network with efficient methods. Therefore, terrorist networks have drawn attention to many researchers all around the world.

Rhodes [14] made statistics for node attributes based on attributes data and computed the factors pairwise attribute values influencing the link between nodes with these attribute values. He chose likelihood ratio in regular Bayesian setting as the similarity measure.

However, methods using topological structure datasets should assume models the network distribution follows firstly. For example, Kim used Kronecker graph model to solve the network completion problem [17], because Leskovec [18] identified that Kronecker graph model can well fit real-world networks with degree distribution, clustering coefficient, etc. Kim chose the model and used expectation maximum method to estimate parameters involved in this model and inferred links with this model. However, there also some networks that follow hierarchical structure [19].

Backstrom and Leskovec [20] proposed a supervised random walk method that used both attribute datasets and topological structure datasets. They built a function to score the true-positive links (predicted links are correct) more weight than false-positive (predicted links are false) according to the observed data. Then, they used this weighted network to supervise random walk method to make inferences.

3 Problem Definition

In the link problems setting, we use an undirected graph $G = (V, E)$ to represent a social network, where edges in E represent interactions between $N = |V|$ nodes in V . We also have categorical attributes for nodes besides network structure. Take open source data from the believed-defunct Greek terrorist group November 17 (N17) [1] as an example; there are 22 terrorists and 64 interactions in this network. In addition to the network structure, each terrorist has common attributes, such as resources (money, weapons, safe houses), role, and faction, although they may take specific values.

We denote the number of distinct attributes M for a specific terrorist network. Attributes of a node u are then represented as a M -dimensional column vector \vec{a}_u with the i^{th} take binary values representing whether a terrorist have it or not, and take categorical values representing which type a terrorist belongs to in terms of this attribute. We denote by $A = [\vec{a}_1, \vec{a}_2, \dots, \vec{a}_N]$ the attribute matrix for all nodes.

Let $P_i = (G_i, A_i)$ and $P_j = (G_j, A_j)$ be snapshots of a terrorist network at times i and j . Then, the link prediction problem involves using G_i and A_i to predict the terrorist structure G_{i+1} . When $i < j$, new links are predicted. When $i > j$, missing links are predicted.

4 Models and Methods

The link prediction problem can be defined as follows; given a snapshot of a social network at time t , we seek to accurately predict the edges that will be added to the network during the interval from time t to a given future time t' and the “missing” links that are not detected at time t . We give some definitions firstly.

4.1 Similarity Measure Definition—Likelihood

We denote some metrics in this similarity measure as follows. Positive links are those links that connect any two individuals in the population, whereas negative links are simply the absence of links. In a Bayesian approach, for a given probability of a link $P(\text{pos})$, the “prior” odds of finding a positive link is given by

$$O_{\text{prior}} = \frac{P(\text{pos})}{1 - P(\text{pos})} = \frac{P(\text{pos})}{P(\text{neg})} \quad (1)$$

However, the “posterior” odds is the odds of finding a positive link after we have considered N pieces of evidence (in our case the attribute) with values $A_1 \dots A_N$ and is given by

$$O_{\text{post}} = \frac{P(\text{pos}|A_1 \dots A_N)}{P(\text{neg}|A_1 \dots A_N)} \quad (2)$$

According to Bayes’ rule, the prior and posterior odds are related by

$$O_{\text{post}} = L(A_1 \dots A_N) O_{\text{prior}} \quad (3)$$

When the pieces of evidence under consideration are conditionally independent, this likelihood ratio can be factorized into a product of individual likelihoods.

$$L(A_1 \dots A_N) = \frac{P(A_1 \dots A_N | \text{pos})}{P(A_1 \dots A_N | \text{neg})} = \prod_{i=1}^N \frac{P(A_i | \text{pos})}{P(A_i | \text{neg})} \quad (4)$$

This similarity proposed by Rhodes [14] is the same as the definition of multiplicative attribute graph model put forward by Kim [21]. They both used the principle of multiplicative independent attribute data to judge the level of pairwise similarities between nodes.

4.2 Sequential Bayesian Updating

The above method relied much on observed data. However, these data are sometimes of insufficient quantity. We hope to find a new algorithm to estimate the statistical confidence of the results a priori and update the similarity measure likelihood. Therefore, we consider the method of sequential Bayesian updating.

We consider data arriving sequentially x_1, \dots, x_n, \dots and wish to update inference on an unknown parameter or state θ . In a Bayesian setting, we have a prior distribution $\pi(\theta)$, and at time n , we have a density for data conditional on θ as,

$$f(x_1, \dots, x_n | \theta) = f(x_1 | \theta) f(x_2 | x_1, \theta) \dots f(x_n | x_{n-1}, \theta) \quad (5)$$

where we have let $x_i = (x_1, \dots, x_i)$. Now we consider a Markovian model for the state dynamics of the form,

$$f(\theta_0) = \pi(\theta_0), f(\theta_{i+1}|\theta_i) = f(\theta_{i+1}|\theta_i) \tag{6}$$

where the evolving states $\theta_0, \theta_1, \dots$ are not directly observed, but information about them is available through sequential observations $X_i = x_i$, where

$$f(x_i|\theta_i, x_{i-1}) = f(x_i|\theta_i) \tag{7}$$

So, the joint density of states and observations is

$$f(x_n, \theta_n) = \pi(\theta_0) \prod_{i=1}^n f(\theta_i|\theta_{i-1})f(x_i|\theta_i) \tag{8}$$

Therefore, the link prediction can be reduced to this problem, which is to predict $f(\theta_{n+1}|x_n)$ depending on the observed density function $f(\theta_n|x_n)$.

In our scenario setting, we adopt a recursive way to estimate the statistical confidence of the results a prior and then regenerate graphs to make inferences. So, θ_0 in this method represents the observed graph, and θ_n represents the new regenerated graph G_n after n times of recursion, while x_n represents the attributes in the graph in the recursion of n times. The new likelihood $L_{n+1}(\theta)$ can be calculated by combining regenerated graph G_n and its corresponding attribute datasets.

If we use the current distribution and the dynamic model,

$$f(\theta_{n+1}|x_n) = \int_{\theta_n} f(\theta_{n+1}|\theta_n)f(\theta_n|x_n)d\theta_n \tag{9}$$

When a new observation $X_{n+1} = x_{n+1}$ is obtained, we can use sequential Bayesian updating to update this distribution,

$$f(\theta_{n+1}|x_{n+1}) \propto f(\theta_{n+1}|x_n)f(x_{n+1}|\theta_{n+1}) \tag{10}$$

5 Algorithm Framework Procedure

We develop an algorithm where Step A is used to make statistics to compute pairwise similarities between attribute values based on the observed link patterns and nodes or edges attribute data.

In the Step B, we consider adopting sequential Bayesian updating to update the structure of the observed graph. Assuming attributes are independent, we use the principle of multiplicative independent attribute data to obtain pairwise similarities between nodes, which are represented by likelihood ratio. We chose the maximum likelihood among all pairs of nodes that are not linked in the graph as the most convinced missing link or the link most likely to occur in the future in Step B.

The algorithm will not stop until it meets the condition in the Step C. The details of this algorithm are shown in Table 1.

Table 1 Link prediction algorithm procedure based on sequential Bayesian updating

Input : Observed network G_0 containing n points, ($n = 17$ in our experiment), positive links denoted by pos, negative links denoted by neg, adjacent matrix $\text{link}[n][n]$, nodes or edges attribute data (each node I has N attributes, $A_{i1}, A_{i2}, \dots, A_{iN}$), and each attribute j has M attribute values.

Output: Inferred network G_m containing m predicted links after m times of recursion.

For step $s = 0$ to m

Step A:

for attribute $i = 1$ to N do

for attribute values $k = 1$ to M do

for attribute values $r = k+1$ to M do

Calculate the likelihood of each pair of values of the i^{th} attribute contained in the network G_s . Likelihood is equal to the prior divided by the posterior in the regular Bayesian settings.

$$L(A_{ik}, A_{ir}) = \frac{P(A_{ik}, A_{ir} | pos)}{P(A_{ik}, A_{ir} | neg)}$$

end for

end for

end for

Step B:

Initialize $L_{\max} = 0$.

for node $v = 1$ to n do

for node $u = v+1$ to n do

Calculate the similarity measure of each pair of nodes according to their nodes or edges attribute data in G_s .

$$L(v, u) = L(A_{v1i}, A_{u1i}) \times \dots \times L(A_{vNi}, A_{uNi})$$

If $L_{\max} \leq L(v, u)$ & $\text{link}[v][u] \neq 1$, then

$$L_{\max} = L(v, u).$$

Return v, u .

Else

Continued.

end for

end for

Step C:

Add the edges between v and u obtained in Step B. Put the edges as predicted link to the network G_s , and then go to Step A.

end for

6 Experiments and Results

As terrorist networks consisting of nodes or edges attribute datasets are hard to obtain, we adopt the datasets of Greek terrorist organization “Revolutionary Organization November 17” (N17) described in the paper of Rhodes [14]. The full topological structure of this organization is shown in Fig. 1.

We randomly delete 50 % extant links and assume this sample graph as observed graph in order to test our algorithm. This sample graph is shown in Fig. 2.

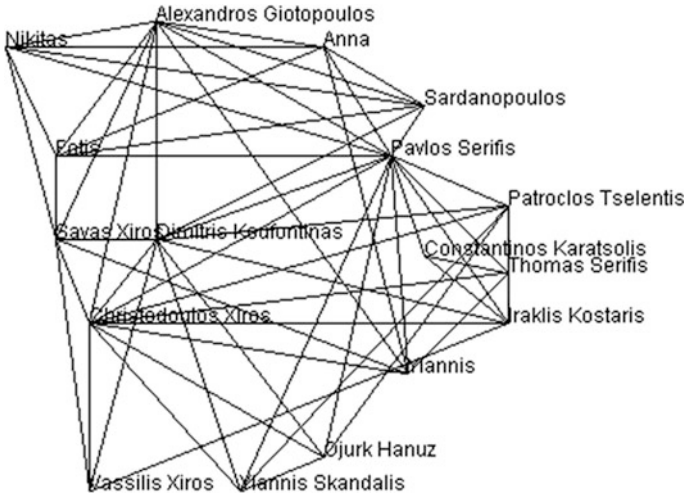


Fig. 1 The full “N17” network from open source reporting

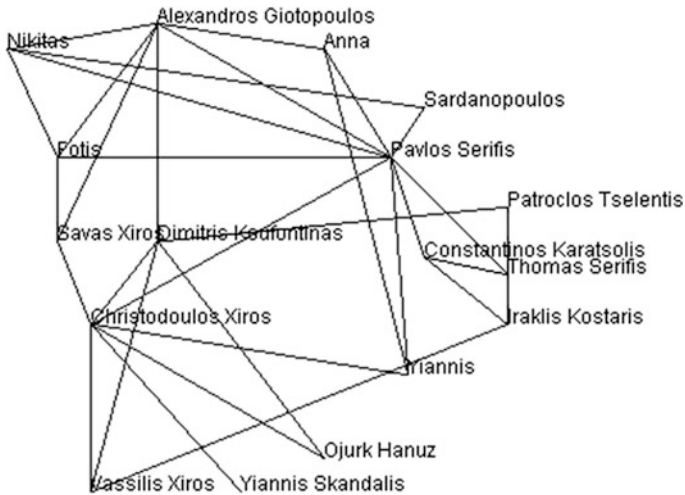


Fig. 2 A sample of the full network generated by removing a random selection of 50 % of the links

Table 2 Attribute data for “N17” network

Name	Attribute			
	Resources control	Role	Faction	Degree
Pavlos Serifis	0	L	S	9
Christodoulos Xiros	2	L	K	7
Savas Xiros	2	O	K	3
Alexandros Giotopoulos	1	L	G	6
Dimitris Koufontinas	3	L	K	5
Anna	0	L	G	3
Iraklis Kostaris	2	O	S	3
Nikitas	0	L	G	4
Patroclus Tselentis	2	O	S	2
Sardanopoulos	1	L	S	2
Fotis	0	L	G	4
Yiannis	2	O	–	3
Thomas Serifis	2	O	S	4
Ojurk Hanuz	2	O	–	2
Yiannis Skandalis	2	O	–	1
Vassilis Xiros	2	O	K	3
Constantinos Karasolis	2	O	S	3
Constantinos Telios ^a	2	O	K	0
Dionysis Georgiadis ^a	2	O	K	0
Elias Gaglias ^a	2	O	K	0
Sotirios Kondylis ^a	2	O	–	0
Vassilis Tzortzatos ^a	2	O	K	0

^a These individuals have not survived the sampling process, and so, we assume that they are not detected in the initial data-gathering operation

We assume counter-terrorism agencies have collected the raw data shown in Fig. 2. Using the equation in the above algorithm, we can calculate the similarity measures $\frac{P(A_{ik}|pos)}{P(A_{i(k+1)}|neg)}$ (A_{ik} and $A_{i(k+1)}$ are different values that attribute A_i can take) between nodes attribute values shown in Fig. 2 and nodes attribute datasets shown in Table 2. In order to judge the similarity measures between nodes, we simply use multiplicative products of those similarities between nodes attribute values to obtain pairwise similarities between nodes. The nodes attribute datasets and topological structure data (degree) of N17 are shown in Table 2.

Using the algorithm procedure described above, we can first get pairwise similarities between attribute values. In this paper, we just list the similarity measures between resources attribute values (the other three attributes can be calculated as the same way), which is shown in Fig. 3.

We then use the second equation in Table 1 to obtain the original similarity measures matrix $L_n(\theta)$, $n = 0$, which is shown in Fig. 4. This is a matrix consisting of 17×17 dimensions that correspond to the total nodes in the graph, and each

ROOURCES	pos links	neg links	P(Resources pos)	P(Resource neg)	Likelihood L
0—0	4	2	0.125	0.019	6.5
0—1	6	2	0.1875	0.019	9.75
0—2	6	34	0.1875	0.327	0.57
0—3	0	4	0	0.038	0
1—1	0	1	0	0.0096	0
1—2	1	19	0.03	0.183	0.17
1—3	1	1	0.03	0.0096	3.25
2—2	10	35	0.31	0.337	0.93
2—3	4	6	0.125	0.058	2.17
3—3	0	0	0	0	Inf
	total=32	total=104			

Fig. 3 Similarity measures for resources attributes

	Pavlos Serifis	Christodoulos Xiros	Savas Xiros	Alexandros Giotopoulos	Dimitris Koufontinas	Anna	Iraklis Kostaris	Nikitas	Patr
Pavlos Serifis	0	0.7193001	0.001	22.52544943	4.95E-07	0.090102	0.007362	45.0509	0.
Christodoulos Xiros		0	0.02	1.87E-05	28.5322359	0.006972	0.001946	2.09E-05	9
Savas Xiros			0	2.90E-04	0.00953597	4.85E-04	3.25E-04	3.88E-04	6
Alexandros Giotopoulos				0	6.58436214	0.51358	3.30E-04	2.054321	1
Dimitris Koufontinas					0	1.78E-05	9.08E-04	2.67E-07	0.
Anna						0	5.54E-04	0.136955	1
Iraklis Kostaris							0	4.43E-04	3
Nikitas								0	6

Fig. 4 The original likelihood matrix between nodes in the observed graph calculated based on their attribute data

value in the matrix represents the likelihood between nodes that lie in the corresponding row and column among the matrix. We just display part of this matrix due to limited space in this paper.

We can infer links that are the most likely missing or the most likely to occur in the future based on the above likelihood matrix. We can see that the likelihood between terrorists Nikitas and Anna is the maximum in that matrix and they are also not linked in the observed graph. Therefore, the first predicted link lies between these two terrorist persons, and it is shown in Fig. 5.

We then make the first recursion with Sequential Bayesian Updating method. We put the above predicted link into the original graph and combine the new link into the original observed graph to update the pairwise similarities between attribute values. So, we recalculate the likelihood matrix $L_n(\theta), n = 1$, which is shown in Fig. 6.

In this way, this prediction process can be continued unless it meets termination conditions. We predict 23 links in order to compare experiments conducted by Rhodes’ algorithm, which also predicted 23 links in total [22]. In this paper, we only display these predicted links every other five steps which are shown in Figs. 7 and 8.

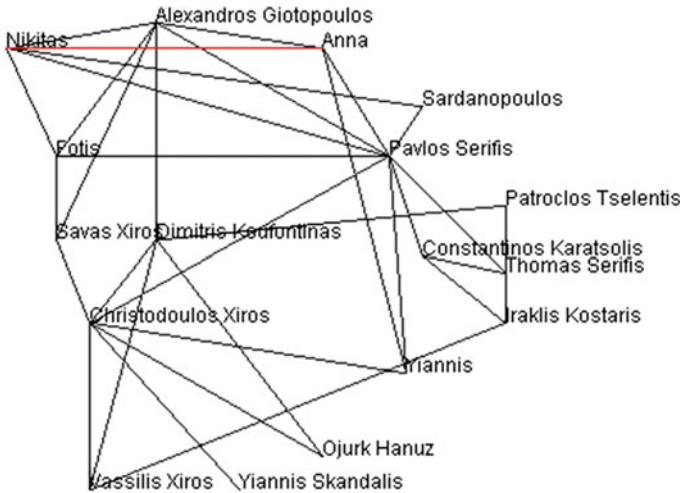


Fig. 5 The first link predicted by our algorithm procedure

Fotis	Yiannis	Thomas Serifis	Ojurk Hanuz	Yiannis Skandalis	Vassilis Xiros	Constantinos Karatsolis	
129.9545	0.012621	3.681166	0.006311	1.14E-04	0.001202	0.007362332	Pavlos S
2.41E-05	0.020434	5.84E-06	0.010217	3.4057046	0.020434	0.001946117	Christod
5.55E-04	0.003414	3.72E-04	6.83E-06	2.05E-05	0.003414	3.25E-04	Savas Xi
5.925926	1.14E-04	0.001321	6.84E-07	2.05E-06	2.90E-04	3.30E-04	Alexandi
3.08E-07	0.009536	1.36E-05	0.09536	4.29E-04	0.009536	9.08E-04	Dimitris
1.410935	1.91E-04	6.33E-04	3.82E-07	1.15E-06	4.85E-04	5.54E-04	Anna
6.33E-04	0.003414	0.002276	6.83E-06	2.05E-05	3.25E-04	0.001991239	Iraklis K
2.469136	2.19E-04	0.001107	2.19E-04	2.29E-06	5.55E-04	6.33E-04	Nikitas
6.33E-04	6.83E-06	0.002276	4.10E-05	4.10E-05	6.50E-07	3.98E-06	Patroclus
0.044556	1.88E-06	6.27E-04	1.13E-05	1.13E-05	1.79E-07	1.10E-06	Sardano

Fig. 6 The first updated likelihood matrix between nodes in the first updated graph calculated based on their attribute data

7 Evaluation

Let us define some performance metrics for predictor evaluation firstly. For each feature vector, a predictor p can make either a positive (P) or a negative (N) prediction concerning the corresponding label. In the positive case, if p is correct, the prediction is said to be true-positive (TP). Otherwise, it is false-positive (FP). It might be useful to define the metric precision as the proportion of TP predictions out of all positive predictions.

$$precision = \frac{|TP|}{|TP + FP|} \tag{11}$$

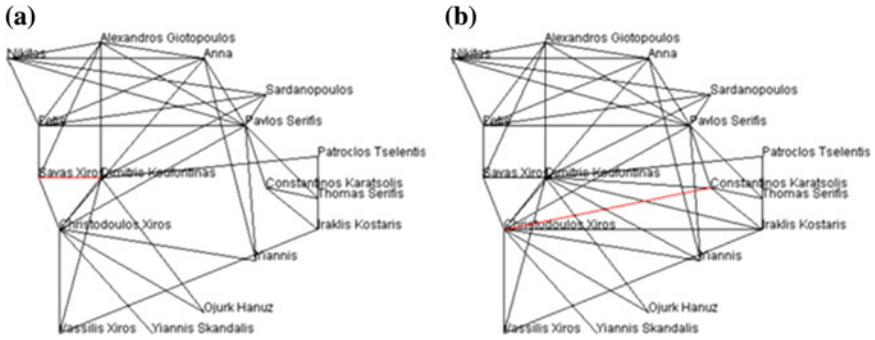


Fig. 7 The a 6th and b 11th links predicted by our algorithm procedure

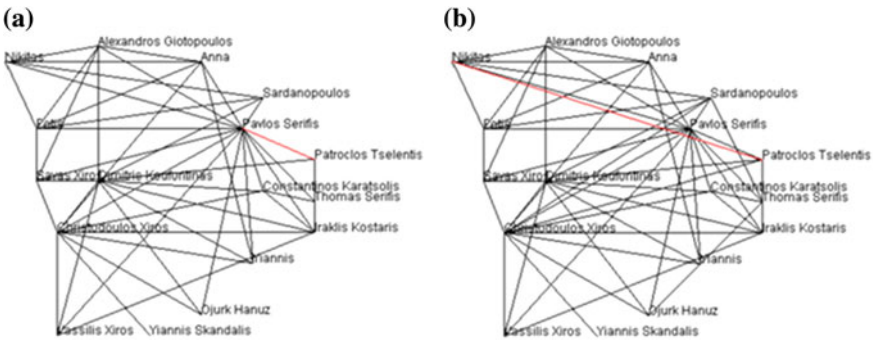


Fig. 8 The a 16th and b 21th links predicted by our algorithm procedure

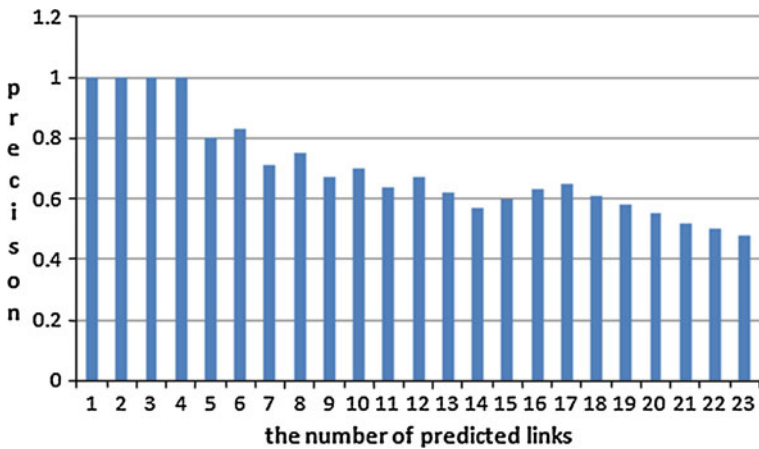


Fig. 9 Precision evaluation on the experimental results

We use this predictor evaluation to test our experimental results, and the performance is shown in Fig. 9.

Of the 23 predictions, 11 are true positives and 12 are false positives, compared to Rhodes' algorithm that 9 are positives and 14 are false positives. As we can see from the above figure, the precision of the experimental results derived from our algorithm has an accuracy of about 48 %, whereas the accuracy of Rhodes' algorithm is 39 % with the same datasets and threshold. The baseline of random connections is calculated by $\frac{32-6}{C_{17}^2} \times 100 \% = \frac{26}{136} \times 100 \% = 19 \%$.

8 Conclusions

We investigated the link prediction problem where the network is incomplete with edges and attribute data missing or being unobserved. We developed an approach based on sequential Bayesian updating combined with Bayesian inference techniques. We compared an approach with Rhodes' algorithm with the same datasets, and the experiment shows that our algorithm performs better by improving accuracy 9 % approximately. This algorithm can work well on small datasets like terrorist networks that have nodes or edges attribute datasets. Therefore, we integrated this algorithm into an emergency decision support system to provide valuable information for counter-terrorism experts.

Acknowledgments This work is supported by National Basic Research Program of China (973 program) with Grant No. 2011CB706900, National Natural Science Foundation of China (Grant No. 70971128), Beijing Natural Science Foundation (Grant No. 9102022), and the President Fund of UCAS (Grant No. O95101HY00).

References

1. Rhodes CJ, Keefe EMJ (2007) Social network topology: a bayesian approach. *J Oper Res Soc* 58:1605–1611
2. Greengrass E (2000) Information retrieval: a survey. University of Maryland, Baltimore
3. Lu L, Medo M, Yeung CH et al (2012) Recommender systems, *Phys Rep* 519(1):1–49
4. Jansen R et al (2003) A Bayesian networks approach for predicting protein–protein interactions from genomic data. *Science* 302:449–453
5. Yu H et al (2008) High-quality binary protein interaction map of the yeast interactome network. *Science* 322:104–110
6. Xiang E (2008) A survey on link prediction models for social network data. PhD thesis, Hong Kong UST, Department of Computer Science and Engineering
7. Resnik P (1995b) Using information content to evaluate semantic similarity in a taxonomy. In: *Proceedings of IJCAI-95*. Montreal, Canada pp 448–453
8. Frakes WB, Baeza-Yates R (1992) Information retrieval: data structures and algorithms. Prentice Hall, US
9. Hindle D (1990) Noun classification from predicate argument structures. In: *Proceedings of the 28th annual meeting of the association for computational linguistics*, pp 268–275

10. Jin EM, Girvan M, Newman MEJ (2001) The structure of growing social networks. *Phys Rev Lett E* 64:046132
11. Salton G, McGill MJ (1983) Introduction to modern information retrieval. McGraw-Hill, Maidenherd
12. Adamic LA, Adar E (2003) Friends and neighbors on the web. *Soc Netw* 25(3):211–230, July 2003
13. Mitzenmacher M (2001) A brief history of log normal and power law distributions. In: Proceedings of the Allerton conference on communication, control, and computing, pp 182–191
14. Rhodes CJ (2009) Inference approaches to constructing covert social network topologies. In: Memon N, Farley JD, Hicks DL, Rosenoorn T (eds) *Mathematical methods in counterterrorism*. Springer, Berlin
15. Xu Z, Tresp V, Yu K, Yu S, Kriegel H-P (2005) Dirichlet enhanced relational learning. In: Proceedings of the 22nd international conference on machine learning. Bonn, Germany, p 1004
16. Yu K, Chu W, Yu S, Tresp V, Xu Z (2007) Stochastic relational models for discriminative link prediction. In: Proceedings of neural information processing systems. MIT Press, Cambridge, MA, pp 1553–1560
17. Kim M, Leskovec J (2011) The network completion problem: inferring missing nodes and edges in networks. In: Proceedings of SDM. pp 47–58
18. Leskovec J, Chakrabarti D, Kleinberg J, Faloutsos C, Ghahramani Z (2010) Kronecker graphs: an approach to modeling networks. *JMLR*
19. Clauset A, Moore C, Newman MEJ (2008) Hierarchical structure and the prediction of missing links in networks. *Nature* 453:98–101
20. Backstrom L, Leskovec J (2011) Supervised random walks: predicting and recommending links in social networks. In: Proceedings of ACM international conference on web search and data mining (WSDM)
21. Kim M, Leskovec J (2010) Multiplicative attribute graph model of real-world networks. arXiv:1009.3499v2
22. Rhodes CJ (2011) The use of open source intelligence in the construction of covert social networks. Counterterrorism and open source intelligence. *Lect Notes Soc Netw (LNSN 2)*. Springer, Wien

Comparison Between Predictive and Predictive Fuzzy Controller for the DC Motor via Network

Abdallah A. Ahmed and Yuanqing Xia

Abstract In the recent years, a networked control system (NCS) has been broadly employed in control systems due to its cost-effective and flexible structure. The major challenge in networked control systems is the warranty system's performance in case of network-induced time delays or data losses. This paper presents a comparison between predictive and predictive fuzzy controller for the DC motor via network. The networked predictive control (NPC) and networked predictive fuzzy control (NPFC) strategy are proposed to compensate the random time delay and data packet dropout in forward channel which can effectively achieve desired control performance of networked control systems. Simulation results are presented to demonstrate the proposed performance.

Keywords Predictive · Predictive fuzzy · Networked control system · DC motor

1 Introduction

The NCSs are composed of the following components: plant, sensors, actuators, and controllers which are coordinated through the communication network. Networked control systems are now extensively used in different industries, ranging from automated manufacturing plants to automotive and aero-spatial applications. This evolution of stand-alone control systems to networked control systems brought many attractive advantages, which include low costs of media and reduce the complexity in wiring connections, simple installation and maintenance,

A. A. Ahmed (✉) · Y. Xia
School of Automation, Beijing Institute of Technology, 100081 Beijing, China
e-mail: abdouahmed12@gmail.com

Y. Xia
e-mail: xia_yuanqing@bit.edu.cn

increased system agility, higher reliability, and greater flexibility [1]. For these reasons, the networked control architecture is already used in many applications, particularly where weight and volume are of consideration, for example in automobiles and aircraft [2, 3]. Meanwhile, the motion control applications can be found in almost every part of industry, from factory automation and robotics to high-tech computer hard disk drives. They are used to regulate mechanical motions in terms of position, velocity, acceleration and/or to coordinate the motions of multiple axes or machine parts. Furthermore, DC motor drives have been extensively used in such applications where the accurate speed tracking is required, and in despite the fact that AC motors are burly, cheaper, and lighter, DC motor is still a very popular choice in particular applications. It is known as a typical plant in the teaching on the control theory and researching. So, it is very meaningful to add the network to DC motor drive control system.

On the other hand, the random network delay in the forward channel in NCS has been studied [4]. So, this paper proposes on networked predictive controller applied at DC motor, and the results are compared with networked predictive fuzzy control. The simulation of the networked predictive fuzzy control is conducted to provide a guide in the experiment.

2 Mathematical Model of DC Motors

Two balance equations can be developed by considering the electrical and mechanical characteristics of the system. Because of the complexity of dynamic system problems, idealizing assumptions will be made. These assumptions are

Assumption 2.1 The brushes are narrow, and commutation is linear.

Assumption 2.2 The armature is assumed to have no effect on the total direct-axis flux because the armature wave is perpendicular to the field axis.

Assumption 2.3 The effects of the magnetic saturation will be neglected.

The electric circuit of the armature and the free body diagram of the rotor is shown in Fig. 1. In a DC motor, the electromagnetic torque is produced by the interaction of the field flux and armature current. A back EMF (speed voltage) is produced when the motor is rotating [5].

The interaction of the T_e with the load torque is given by:

$$T_e = J \frac{d\omega}{dt} + B\omega + T_l \quad (1)$$

From Fig. 1, we can write

$$V_t = R_a i_a + L_a \frac{di_a}{dt} + E_a \quad (2)$$

where T_l is load torque, K_T is torque constant, and K_b is back EMF constant. ω and θ are angular speed and angular displacement, respectively.

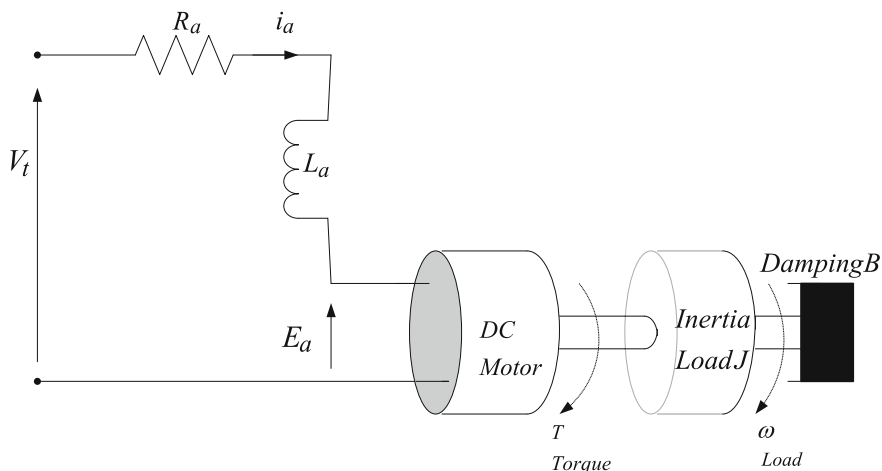


Fig. 1 DC motor equivalent circuit

V_t , R_a , and L_a are motor terminal voltage, armature resistance, and armature inductance, respectively.

J and B are moment of inertia and friction coefficient, respectively.

By choosing i_a , ω and θ as the state variables and V_t as input, the output is chosen to be ω , the state space model of DC motors is given as follows:

$$\begin{bmatrix} \dot{i}_a \\ \dot{\omega} \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} \frac{-R_a}{L_a} & \frac{-K_b}{L_a} & 0 \\ \frac{K_T}{J} & \frac{-B}{J} & 0 \\ 0 & 1 & 0 \end{bmatrix} x + \begin{bmatrix} \frac{1}{L_a} \\ 0 \\ 0 \end{bmatrix} V_t. \tag{3}$$

$$y = [0 \quad 1 \quad 0] \begin{bmatrix} i_a \\ \omega \\ \theta \end{bmatrix}$$

where

$$x = [i_a \quad \omega \quad \theta]^T$$

are state variables.

By defining system matrices A_c , B_c and C_c as follows:

$$A_c = \begin{bmatrix} \frac{-R_a}{L_a} & \frac{-K_b}{L_a} & 0 \\ \frac{K_T}{J} & \frac{-B}{J} & 0 \\ 0 & 1 & 0 \end{bmatrix}, \tag{4}$$

$$B_c = \begin{bmatrix} \frac{1}{L_a} \\ 0 \\ 0 \end{bmatrix}, C_c = [0 \quad 1 \quad 0], \quad (5)$$

we can obtain discrete state space of DC motors dynamic with step time T_s as follows:

$$\begin{cases} x_{k+1} &= Ax_k + Bu_k \\ y_k &= Cx_k \end{cases} \quad (6)$$

3 Networked Predictive Controller for DC Motor with Network Delay

Model predictive control (MPC) is accepted control strategy based on using a model to predict at each sampling period the future evolution of the system from the current state along a given prediction horizon [6, 7]. The best control sequence of control inputs is obtained by minimizing an optimization criterion; the first control signal is applied to the process, and the procedure is repeated at the next sampling instant. This characteristic makes the MPC approach very suitable to integrate the input/output constraints into the online optimization problem as well as to compensate time delays, which increases the possibility of its application in the synthesis and analysis of NCSs.

In a NCS, the sensors have the duty of measuring the outputs of the plant and sending the samples to the controller through the network. The controller receives the measurements from the sensors, then calculates the control command, and sends the values through the network to the actuators. The actuators have the task of applying the control commands received through the network to the physical plant [8]. The structure of the networked predictive control strategy for DC motor is shown in Fig. 2. The controller and the plant are connected by the network, which causes random communication time delay and data packet dropout. Traditional control method cannot well compensate for these two reasons. However, one of the features of the networked control systems is that a set of control sequences can be packed and transmitted from one location to another location at the same time through a network channel. This feature provides a possibility to compensate for the network time delay and the network data packet losses by transmitting a set of future control sequences at the current time. For more details about the network-induced time delay, see [4]. In this paper, we consider the case in which the system controller is far away from the plant but the sensor is near to the plant. So, the network delay in the feedback channel is not considered. A networked predictive control with random network delay in the forward channel is proposed. The main part of the scheme is the networked predictive controller, to

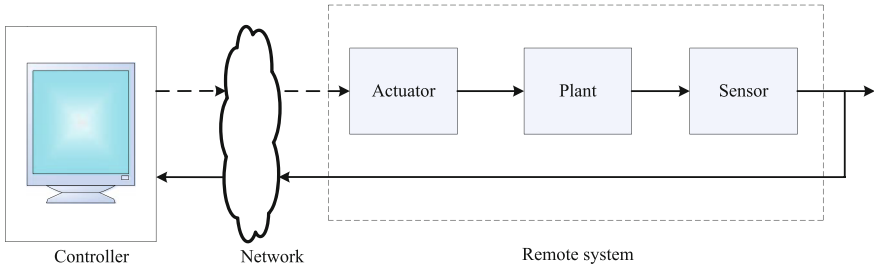


Fig. 2 Network control system diagram

compensate network delay in the forward channel for DC motor and achieve the desired control performance.

The plant studied in this paper is described by Eq. (6). The state observer is designed as

$$\hat{x}_{t+1|t} = A\hat{x}_{t|t-1} + Bu_t + L(y_t - CA\hat{x}_{t|t-1}) \tag{7}$$

where $\hat{x}_{t+1|t} \in R^3$ and $u_t \in R^1$ are the one-step ahead state prediction and the input of the observer at time t , respectively. $y_t - CA\hat{x}_{t|t-1}$ is innovation. The matrix $L \in R^{n \times l}$ can be obtained using observer design approaches. Following the state observer described by Eq. (7), based on the output data up to t , the state prediction from time $t + 1$ to N are constructed as

$$\begin{aligned} \hat{x}_{t+1|t} &= A\hat{x}_{t|t-1} + Bu_t + L(y_t - CA\hat{x}_{t|t-1}) \\ \hat{x}_{t+2|t} &= A\hat{x}_{t+1|t} + Bu_{t+1|t} \\ &\vdots \\ \hat{x}_{t+N|t} &= A\hat{x}_{t+N-1|t} + Bu_{t+N-1|t} \end{aligned} \tag{8}$$

where integer N is defined as $N \triangleq N_1 + N_2$, N_1 is maximum time delay in the forward channel, and the integer N_2 denotes the maximum of number of consecutive data packet dropout.

In order to compensate the network transmission delay, a network delay compensator is proposed. A very important feature of the network is that it can transmit a set of data at the same time. So, it is assumed that predictive control sequence at time t is packed and sent to the plant side through a network. The network delay compensator chooses the latest control value from the control prediction sequences available on the plant side.

Assume that the controller is of the following form:

$$u_t = K\hat{x}_{t+1|t} + K_i \int e_k \tag{9}$$

where $K \in R^{1 \times 3}$ is the state feedback control matrix to be determined using modern control theory, and K_i is a tuning gain for integration of error. Based on this controller, we can have predictive controller as follows

$$\hat{u}_{i|t} = K(\omega_r - \hat{x}_{i|t}^2) + K_i \sum_{k=0}^t e_k \tag{10}$$

where ω_r is speed reference, and $\hat{x}_{i|t}^2$ denotes the second element of vector $\hat{x}_{i|t}$, i.e., the predictive speed at time t .

4 Networked Predictive Fuzzy Controller for DC Motor with Network Delay

We consider the same case mentioned in Sect. 3 where the controller (predictive fuzzy) at the local node is far away from the plant and the manipulated variables are transmitted through networks. The structure of the networked predictive fuzzy control strategy for DC motor is shown in Fig. 3. In this section, in order to compensate for the random time delay and data packet dropout in forward channel for DC motor, a networked predictive fuzzy control (NPFC) strategy is proposed. The strategy mainly consists of three parts: a model predictor, a fuzzy controller, and a network delay compensator. The model predictor is used to generate a series of predictive outputs according to the past control actions and the past plant outputs. The fuzzy controller is designed to produce a set of future control predictions with the error between the reference and the predictive outputs from the model predictor. The network delay compensator is used to compensate for the unknown random delay and the data packet losses by selecting appropriate control sequences.

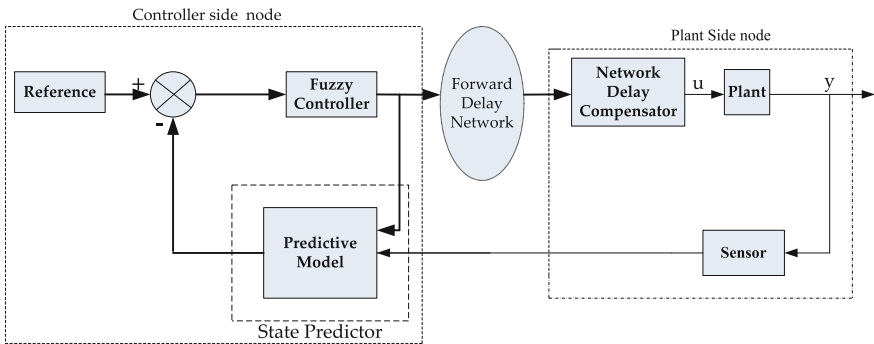


Fig. 3 Networked predictive fuzzy control for DC motor

4.1 Fuzzy Logic Control

The purpose of a fuzzy controller is to convert linguistic control rules based on expert knowledge into control strategy [9]. The effective and efficient control using fuzzy logic has appeared as a tool to deal with unsure, imprecise, or qualitative decision-making problems [10, 11]. The FLC involves four stages namely fuzzification, rule-base, inference engine, and defuzzification. The Sugeno-type controller is performed for present control because it has singleton membership in the output variable. Moreover, it can be easily achieved and number of calculations can be reduced [12].

Fuzzification In this work, the motor variables considered are speed ω and current i_a . The speed ω is the control object of FLC. Let ω_r denote the reference speed, then the definitions for error e_k and change in error Δe_k are given in (11) and (12).

$$e_k = \omega_{r,k} - \omega_k \quad (11)$$

$$\Delta e_k = e_k - e_{k-1} \quad (12)$$

Five linguistic variables are utilized for fuzzifying the input variable e_k and Δe_k are as follows, Negative Big (NB), Negative Small (NS), Zero (Z), Positive Small (PS), and Positive Big (PB). There are various types of membership functions, such as triangular shaped, Gaussian, sigmoidal, pi shaped trapezoidal shaped, bell shaped. the triangular membership function is used for simplicity and also to reduce the calculations [13]. Usually seven membership functions are favored for accurate result. In this work, only five membership functions are utilized for the input, i.e., error and alteration in error. In order to minimize the number of membership function, the width of the membership functions is kept different.

Defuzzification The linguistic variables are transformed into a numerical variable [14]. As the weighted sum method is considered to be the best well-known defuzzification method, it is utilized in the present model. The defuzzified output is the duty cycle dc_k . The change in duty cycle Δdc_k can be obtained by adding the pervious duty cycle pdc_k with the duty cycle dc_k which is given in Eq. (13).

$$\Delta dc_k = dc_k + pdc_k \quad (13)$$

Rule table and Inference Engine The general rule can be written as “If $e(k)$ is X and $\Delta e(k)$ is Y then $\Delta dc(k)$ is Z ,” where X, Y , and Z are the fuzzy variable for $e(k)$, $\Delta e(k)$ and $\Delta dc(k)$, respectively .

4.2 Networked Predictive Fuzzy Controller

As we mentioned above, in order to compensate for the random time delay and data packet dropout in forward channel for DC motor, a networked predictive fuzzy control (NPFC) strategy is proposed.

The state observer is designed as Eq. (8) mentioned at Sect. 3.

Now assume that the controller is of the following form:

$$u = \text{fuzzy} (e, \Delta e) + K_i \int e_k \quad (14)$$

where symbol (fuzzy) denotes fuzzy controller operation mentioned at Sect. 4.1, and K_i is a tuning gain for integration of error. Based on this controller, we can have predictive fuzzy controller as follows

$$\hat{u}_{i|t} = \text{fuzzy} (\omega_r - \hat{x}_{i|t}^2, \hat{x}_{i|t}^2 - \hat{x}_{i-1|t}^2) + K_i \sum_{k=0}^t e_k \quad (15)$$

where ω_r is speed reference, and $\hat{x}_{i|t}^2$ denotes the second element of vector $\hat{x}_{i|t}$, i.e., the predictive speed at time t .

At every step, the controller sends a set of predictive manipulated variables in a data packet:

$$\{\hat{u}_{t+i|t} \mid i = 0, 1, \dots, N\}$$

while at the remote node, a compensator is designed to choose a predictive manipulated variables to plant as the actual manipulated input from its receiving buffer:

$$u_t = \hat{u}_{t|t-i} \quad (16)$$

From the above, it is shown that in the case of no network delay in the communication channel, the input to the plant actuator is the output of the controller. In the case of a delay iT , where T is the sampling period, the control input to the actuator is the i th-step ahead control prediction received in the current sampling period.

5 Simulation Results and Discussion

In this section, a comparison of proposed model has been simulated using Matlab, and the designed networked predictive controller (NPC) and networked predictive fuzzy controller (NPFC) are tested. The sampling interval is $T = 0.01$. The parameters of the studied motor used for simulation are given in Table 1.

Table 1 The parameters of the studied motor

DC motor parameters	Values
Motor rating	5HP
DC supply voltage	220 V
Motor rated current	4.3 A
Armature resistance R_a	0.6
Armature inductance L_a	0.008 H
Inertia constant J	0.011 N – m ²
Damping constant B	0.004 Nm/rad/s
Back EMF constant K_b	0.55
Torque constant K	0.55

Time delay exist in the forward communication channel, which can be seen from Fig. 4. The delay time for comparison is assumed to be 0 (no delay time), 1, 4, 8, 12, and 20 ms. The network data packet losses occur due to longer time delays. However, the network predictive control like the network predictive fuzzy control can effectively compensate the network time delay and the network data packet losses with shorter time delays, while for the longer time delays the networked predictive fuzzy control works effectively when compared to the networked predictive control. These can be seen from the comparison between the networked predictive control and the networked predictive fuzzy control as shown in Figs. (5, 6, 7). Furthermore, it is also observed that when the time delays are longer, the system response is oscillated.

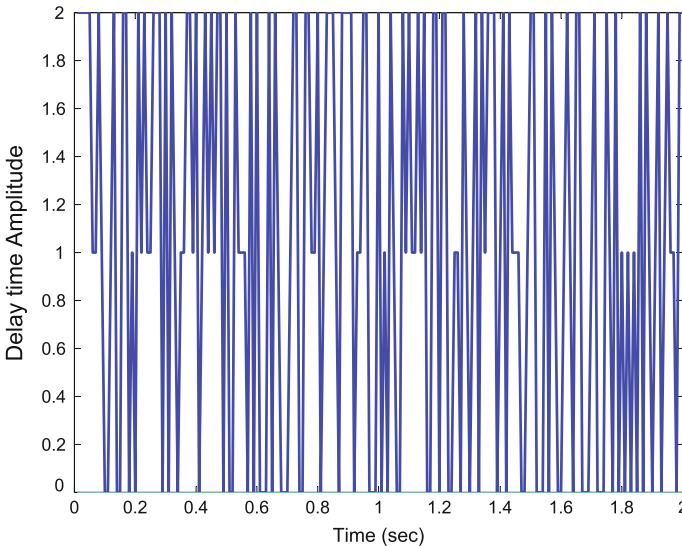


Fig. 4 Forward delay in the network

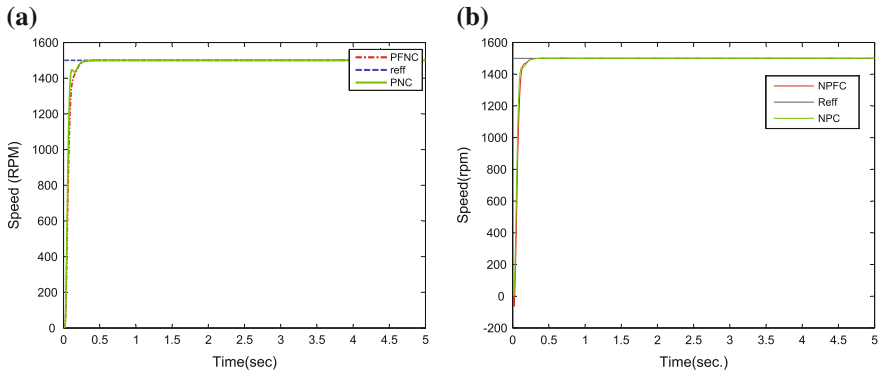


Fig. 5 **a** Comparison of NPFC with 0 ms. Delay time and NPC. **b** Comparison of NPFC with 1ms. Delay time and NPC

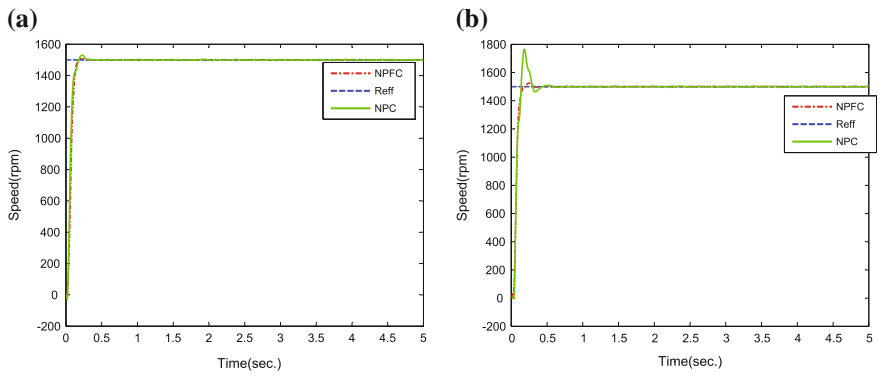


Fig. 6 **a** Comparison of NPFC with 4 ms. Delay time and NPC. **b** Comparison of NPFC with 8 ms. Delay time and NPC

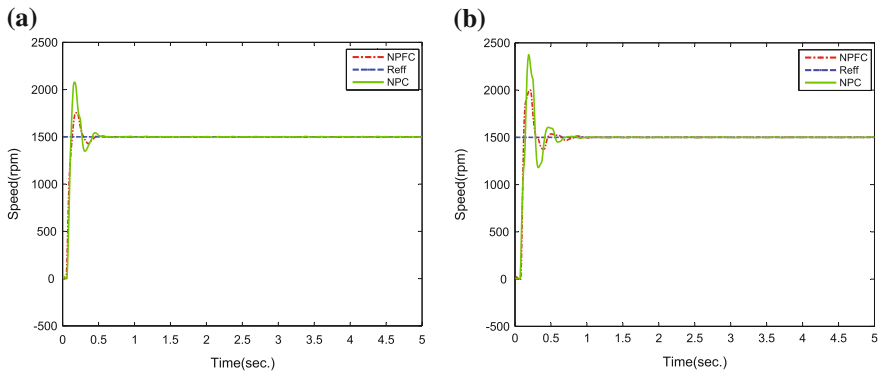


Fig. 7 **a** Comparison of NPFC with 12 ms. Delay time and NPC. **b** Comparison of NPFC with 20 ms. Delay time and NPC

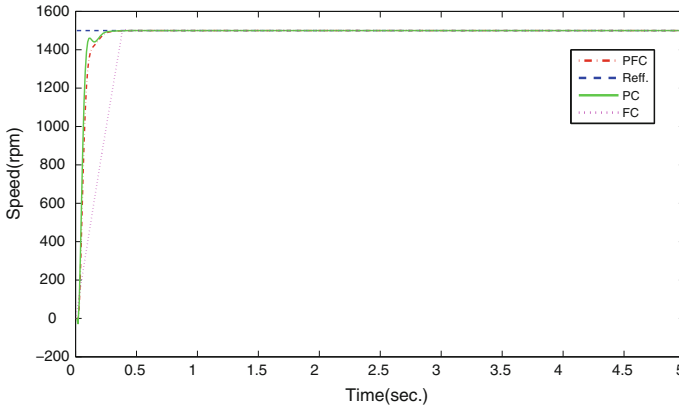


Fig. 8 Comparison of PFC, PC, and FC with speed reference 1,500 rpm

In order to evaluate and show difference between three systems, comparison between the FC, PC, and PFC controller are shown in the Fig. 8.

6 Conclusion

Networks and their applications play a promising role for real-time performance networked control in industrial applications. The major concerns are the network-induced delays and data losses that are provided by the network which affects the performance of the networked control systems. This paper has compared the predictive fuzzy logic controller and predictive fuzzy controller in a networked for DC motor. The results show that the performance of networked control DC motor has obtained better results by using predictive fuzzy logic controller than predictive fuzzy controller.

Acknowledgments This work was supported by the National Basic Research Program of China (973 Program) (2012CB720000), the National Natural Science Foundation of China (61225015, 60974011), the PhD Programs Foundation of Ministry of Education of China (20091101110023, 20111101110012), and Beijing Municipal Natural Science Foundation (4102053, 4101001).

References

1. Johansson KH, Törngren M, Nielsen L (2005) Vehicle applications of controller area network. In: Handbook of networked and embedded control systems, vol 1(13). Springer
2. Xia Y, Chen J, Liu GP, Rees D (2007) Stability analysis of networked predictive control systems with random network delay. In: Proceedings of the IEEE international conference network sensor control. London, UK, pp 815–820.

3. Liu X, Liu Y, Mahmoud MS, Deng Z (2010) Modeling and stabilization of MIMO networked control systems with network constraints. *Int J Innov Comput Inf Control* 6(10):4409–4420
4. Xia Y, Fu M, Liu B, Liu GP (2009) Design and performance analysis of networked control systems with random delay. *J Syst Eng Electron* 20(4):807–822
5. Fitzgerald AE, Kingsley C Jr, Kusko A (1971) *Electric machinery*, 3rd edn. USA.
6. Camacho EF, Bordons C (2004) *Model predictive control*. Springer
7. Maciejowski JM (2002) *Predictive control with constraints*. Prentice Hall, Harlow
8. Onat A, Parlakay EM (2007) Implementation of model based networked predictive control system. 9th Real-Time Workshop (RTLW09), 2(13):85–93. Linz, Austria
9. Zhang BS, Edumunds JM (1991) On fuzzy logic controllers. In: IEE international conference on control. Edinburg, UK, pp 961–965
10. Soliman HF, Mansour MM, Kandil SA, Sharaf AM (Sep 1995) A robust tunable fuzzy logic control scheme for speed regulation of DC series motor drives. *Electr Comput Eng Can Conf* 1(5–8):296–299
11. Majed J, Houda C, Houssein J, Braiek B (2008) Naecur Fuzzy logic parameter estimation of an electrical systems. *Signals and devices, IEEE SSD*. In: 5th international multi-conference, pp 1–6
12. Senthil Kumar N, Sadasivam V, Muruganandam M (Aug 2007) A Low-cost four-quadrant chopper-fed embedded DC drive using fuzzy controller. *Int J Electr Power Compon Syst* 35(8):907–920
13. El-kholy AE, Dabroom AM (2006) Adaptive fuzzy logic controllers for DC drives. A survey of the state of the art. *J Electr Syst*:116–145
14. Yousef HA, Khalil HM (Jul 1995) A fuzzy logic-based control of series DC motor drives. *Proc IEEE Int Symp* 2(10–14):517–522

Trajectory Tracking Method for UAV Based on Intelligent Adaptive Control and Dynamic Inversion

Juan Dai and Yuanqing Xia

Abstract This paper discusses the flight control strategy based on intelligent adaptive control and dynamic inversion. The primary use of the Optimal Control Modification (OCM) adaptive control is to add damping to the neural network controller weight update law so as to reduce high-frequency oscillations in the weights and to prevent parameter drift in the absence of persistent excitation. The OCM is applied to the inner loop control of an UAV during flight conditions to investigate its flight control augmentation capability to a dynamic inversion (DI) controller subject to off-nominal flight conditions.

Keywords Intelligent adaptive flight control · Dynamic inversion · UAV · Flight control.

1 Introduction

In recent years, there have been a great many researches on control designs for UAVs with highly nonlinear characteristics using nonlinear control techniques [1–6]. With the great advancements in the microelectronics and precise navigation systems, fully or partially automated UAVs are being used and developed for

This work was supported by the National Basic Research Program of China (973 Program) (2012CB720000), the National Natural Science Foundation of China (61225015,60974011), the PhD Programs Foundation of Ministry of Education of China (20091101110023, 20111101110012), and Beijing Municipal Natural Science Foundation (4102053, 4101001).

J. Dai (✉) · Y. Xia

School of Automation, Beijing Institute of Technology, Beijing 100081, China
e-mail: juandai2011@gmail.com

Y. Xia

e-mail: xia_yuanqing@bit.edu.cn

various missions. Combination of two technologies has led to an enormous rise of interest in tracking and trajectory control for wide range of platforms in both military and civil aviation including small unmanned vehicles, unmanned helicopters, and transport-class aircrafts in the context of next generation air transport [7–11]. So, UAVs are more and more frequently required to follow a pre-designated trajectory.

Adaptive flight controls tend to get more attention and appreciation in air vehicle applications than in space system applications [12, 13]. This is simply due to its fundamental difference in both flight control bandwidth (i.e., fast vs slow) and the mission performance characteristics (i.e., manned aircraft fly numerous missions while spacecraft are unmanned and fly relatively fewer missions). As a result, adaptive flight control design methodologies developed for aircraft systems tend to be more mature than those for space systems and its applications to real platforms like F-18, UAVs.

Adaptive control is a promising technology that could improve closed-loop vehicle dynamics under failure. The ability to accommodate system uncertainties and to improve fault tolerance of a flight control system is a major selling point of adaptive control. There has been a steady increase in the number of adaptive control applications in a wide range of settings such as aerospace, robotics, process control. Adaptive control continues to enjoy the attention and broad support from government agencies, industry, and academia.

The flight trajectory tracking control system is a rather important and complex problem. Because the flight vehicle dynamics demonstrate the character of strong coupling and high nonlinear, the conventional gain scheduling method is insufficiently efficient for precise trajectory tracking. The advent of modern control theories has afforded flight trajectory tracking control lately years [14–21].

The adaptive OCM/NN-based flight control design technique considered in this paper as a control assistant to the DI controller to perform rate stabilization. It is premature at this stage to conclude that the OCM/NN controller is the best control augmentor among its adaptive design counterparts; nevertheless, its inherent ability to interact with the UAVs primary controller (DI controller in this case) in order to maintain the vehicle stability while the baseline fails to stabilize deserves more attention in the control community.

The focus of this paper is to evaluate the OCM control law for its ability to assist the baseline DI controller to maintain the vehicle stability subject to loss of control surfaces. The OCM adaptive law is implemented as an augmenting controller to the existing DI controller.

The paper is organized as follows. A nonlinear UAV model is derived in Sect. 29.2. The UAV trajectory tracking control scheme based on intelligent adaptive control and dynamic inversion is designed in Sect. 29.3, and an intelligent adaptive controller and dynamic inversion controller are effectively combined. The stability analysis of the controller is presented in Sect. 29.4. The paper ends with the conclusion in Sect. 29.5.

2 Nonlinear UAVs Model

A number of assumptions have to be made before proceeding with the derivation of the equations of motion:

1. The aircraft is a rigid body.
2. The earth is flat and nonrotating and regarded as an inertial reference.
3. The mass is constant during the time intervals over which the motion is considered.
4. The mass distribution of the aircraft is symmetric.

The state vector used to model the nonlinear aircraft dynamics is given by

$$x = [\phi \ \theta \ \psi \ \alpha \ \beta \ Vp \ q \ r]^T \quad (1)$$

where ϕ , θ , and ψ are Euler's roll, pitch, and yaw attitude angles, α and β are the angles of attack and sideslip, V is the total velocity of the aircrafts center of gravity, and p , q , and r are the aircrafts body roll, pitch, and yaw rates, respectively. The control variables are the deflections δ_e , δ_a , and δ_r of the conventional elevator, aileron, and rudder control surfaces, in addition to the throttle setting ξ . Hence, the control vector is given by

$$u = [\delta_e \ \delta_a \ \delta_r \ \xi]^T \quad (2)$$

The thrust T generated by the engine is

$$T = \xi T_{\max}$$

where $\xi \in [0, 1]$ denotes the instantaneous throttle setting, and T_{\max} is the maximum available thrust and is assumed to be a constant.

2.1 State Vector Decomposition

We partition the state vector x into

$$x = [x_1^T; x_2^T; x_3^T]^T = [\phi \ \theta \ \psi; \alpha \ \beta; Vp \ q \ r]^T \quad (3)$$

where $x_1 \in R^3$ is the vector of unactuated states, namely the states with relative degree two (Euler's angles), $x_2 \in R^2$ and $x_3 \in R^4$ are respectively the outer actuated and inner actuated state vectors, both have relative degree one, such that x_3 and u have the same dimension. Hence,

$$x_1 = [\phi \ \theta \ \psi]^T, \quad x_2 = [\alpha \ \beta]^T, \quad x_3 = [Vp \ q \ r]^T \quad (4)$$

2.2 Unactuated Subsystem

Euler's angular kinematical equations of motion for the aircraft are given by

$$\dot{\phi} = p + q \tan \theta \sin \phi + r \tan \theta \cos \phi \quad (5)$$

$$\dot{\theta} = q \cos \phi - r \sin \phi \quad (6)$$

$$\dot{\psi} = q \sin \phi / \cos \theta + r \cos \phi / \cos \theta \quad (7)$$

Hence, the unactuated subsystem is obtained from the above kinematical equations as

$$\dot{x}_1 = A_1(x_1)x_3 \quad (8)$$

where

$$A_1(x_1) = \begin{bmatrix} 0 & 1 & \tan \theta \sin \phi & \tan \theta \cos \phi \\ 0 & 0 & \cos \phi & -\sin \phi \\ 0 & 0 & \sin \phi / \cos \theta & \cos \phi / \cos \theta \end{bmatrix} \quad (9)$$

2.3 Outer Actuated Subsystem

The rates of change of the aerodynamic angles α and β are given by the expressions

$$\begin{aligned} \dot{\alpha} = & q - (p \cos \alpha + r \sin \alpha) \tan \beta + \frac{g}{V \cos \beta} (\cos \alpha \cos \phi \cos \theta + \sin \alpha \sin \theta) \\ & - \frac{1}{mV \cos \beta} [\bar{q} S C_L + T \sin (\alpha + \delta)] \end{aligned} \quad (10)$$

$$\begin{aligned} \dot{\beta} = & p \sin \alpha - r \cos \alpha + \frac{g}{V} (\cos \beta \cos \theta \sin \phi - \cos \phi \cos \theta \sin \alpha \sin \beta + \cos \alpha \sin \beta \sin \theta) \\ & - \frac{1}{mV} [\bar{q} S C_s + T \cos (\alpha + \delta) \sin \beta] \end{aligned} \quad (11)$$

where δ is a constant thrust deflection angle, S is the wing area, the dynamic pressure \bar{q} is defined in terms of the air density ρ and total air velocity V as

$$\bar{q} = \frac{1}{2} \rho V^2 \quad (12)$$

and the dimensionless lift and side force coefficients C_L and C_S are given by

$$C_L = C_{L_0} + C_{L_\alpha} \alpha + C_{L_{\alpha^2}} (\alpha - \alpha_{\text{ref}})^2 + C_{L_q} \frac{\bar{c}}{2V} q + C_{L_{\delta_e}} \delta_e \quad (13)$$

$$C_S = C_{S_0} + C_{S_\beta} \beta + C_{S_{\delta_r}} \delta_r \quad (14)$$

The terms α_{ref} and \bar{c} are constants representing the critical angle-of-attack and the wing mean chord length, respectively. The aerodynamic and control coefficients in the above expressions of C_L and C_S are assumed to be constants around nominal flight operating points.

Hence, the following control vector-explicit state-space subsystem for the outer actuated dynamics

$$\dot{x}_2 = A_2(x_1, x_2, x_3) + B_2(x_2, \bar{q})u \quad (15)$$

where the vector function $A_2(x_1, x_2, x_3)$ and $B_2(x_2, \bar{q})$ are given by

$$A_2 = \begin{bmatrix} q - (p \cos \alpha + r \sin \alpha) \tan \beta + \frac{g}{V \cos \beta} (\cos \alpha \cos \phi \cos \theta + \sin \alpha \sin \theta) - \\ \frac{\bar{q}S}{mV \cos \beta} (C_{L_0} + C_{L_\alpha} \alpha + C_{L_{\alpha^2}} (\alpha - \alpha_{\text{ref}})^2 + C_{L_q} \frac{\bar{c}}{2V} q) \\ p \sin \alpha - r \cos \alpha + \frac{g}{V} (\cos \beta \cos \theta \sin \phi - \cos \phi \cos \theta \sin \alpha \sin \beta \\ + \cos \alpha \sin \beta \sin \theta) - \frac{\bar{q}S}{mV} (C_{S_0} + C_{S_\beta} \beta) \end{bmatrix} \quad (16)$$

$$B_2 = \begin{bmatrix} \frac{-\bar{q}S C_{L_{\delta_r}}}{mV \cos \beta} & 0 & 0 & -\frac{T_{\max} \sin(\alpha + \delta)}{mV \cos \beta} \\ 0 & 0 & \frac{-\bar{q}S C_{L_{\delta_r}}}{mV} & -\frac{T_{\max} \cos(\alpha + \delta)}{mV \sin \beta} \end{bmatrix} \quad (17)$$

2.4 Inner Actuated Subsystem

The rate of change of the total velocity V is given by the expression

$$\begin{aligned} \dot{V} = & g \cos \theta \sin \beta \sin \phi + g \cos \beta (\cos \phi \cos \theta \cos \alpha - \cos \alpha \sin \theta) - \frac{\bar{q}S}{m} C_D \\ & + \frac{1}{m} \cos(\alpha + \delta) \cos \beta T_{\max} \zeta \end{aligned} \quad (18)$$

where the dimensionless drag coefficient C_D is given by

$$C_D = C_{D_0} + C_{D_{\alpha^2}} \alpha^2 \quad (19)$$

The aerodynamic coefficients C_{D_0} and $C_{D_{\alpha^2}}$ are in the above nominal flight operating points.

Hence, \dot{V} is written as

$$\begin{aligned} \dot{V} = & g \cos \theta \sin \beta \sin \phi + g \cos \beta (\cos \phi \cos \theta \cos \alpha - \cos \alpha \sin \theta) - \frac{\bar{q}S}{m} (C_{D_0} + C_{D_{z_2}} \alpha^2) \\ & + \frac{1}{m} \cos(\alpha + \delta) \cos \beta T_{\max} \xi \end{aligned} \quad (20)$$

Euler's system of dynamical equations for angular motion is given by

$$\dot{\omega} = R^{-1} S^\times(\omega) R \omega + R^{-1} \tilde{T} \quad (21)$$

where $\omega = [p \ q \ r]^T$, R is the constant body's inertia matrix about the center of mass of the aircraft, $S^\times(\omega)$ is the cross-product matrix corresponding to the angular velocity ω , the total external moment vector \tilde{T} on the aircraft in the body frame is

$$\tilde{T} = [L \ M \ N]^T \quad (22)$$

The aerodynamic moments L , M , and N are given by

$$L = \bar{q} S b (C_{L_0} + C_{L_\beta} \beta + C_{L_p} \frac{b}{2V} p + C_{L_r} \frac{b}{2V} r + C_{L_{\delta_a}} \delta_a + C_{L_{\delta_r}} \delta_r) \quad (23)$$

$$M = \bar{q} S \bar{c} (C_{M_0} + C_{M_\alpha} \alpha + C_{M_q} \frac{\bar{c}}{2V} q + C_{M_{\delta_e}} \delta_e) + T_{\max} \Delta z \zeta \quad (24)$$

$$N = \bar{q} S b (C_{N_0} + C_{N_\beta} \beta + C_{N_p} \frac{b}{2V} p + C_{N_r} \frac{b}{2V} r + C_{N_{\delta_a}} \delta_a + C_{N_{\delta_r}} \delta_r) \quad (25)$$

where Δz is the moment arm of the thrust vector about the body y -axis, and b is the wing span.

The aerodynamic and control coefficients in the above three expressions of L , M , and N are assumed to be constants around nominal flight operating points.

Hence, $\dot{\omega}$ is written in the following form:

$$\dot{\omega} = R^{-1} S^\times(\omega) R \omega + \bar{q} S R^{-1} [f(x_2, x_3) + B_\omega u] \quad (26)$$

where the vector function $f: R^2 \times R^2 \rightarrow R^3$ is given by

$$f = \begin{bmatrix} b(C_{L_0} + C_{L_\beta} \beta + C_{L_p} \frac{b}{2V} p + C_{L_r} \frac{b}{2V} r) \\ \bar{c}(C_{M_0} + C_{M_\alpha} \alpha + C_{M_q} \frac{\bar{c}}{2V} q) \\ b(C_{N_0} + C_{N_\beta} \beta + C_{N_p} \frac{b}{2V} p + C_{N_r} \frac{b}{2V} r) \end{bmatrix} \quad (27)$$

and $B_\omega \in R^{3 \times 4}$ is the input matrix

$$B_\omega = \begin{bmatrix} 0 & bC_{L\delta_a} & bC_{L\delta_r} & 0 \\ \bar{c}C_{M\delta_e} & 0 & 0 & \frac{T_{\max}\Delta z}{\bar{q}S} \\ 0 & bC_{N\delta_a} & bC_{N\delta_r} & 0 \end{bmatrix} \quad (28)$$

Hence, the aircraft inner actuated dynamics is described by the following control vector-explicit state-space subsystem

$$\dot{x}_3 = A_3(x_1, x_2, x_3) + B_3(x_2, \bar{q})u \quad (29)$$

where the vector function $A_3 : R^3 \times R^2 \times R^4 \rightarrow R^4$ is given by

$$A_3(x_1, x_2, x_3) = \begin{bmatrix} g \cos \theta \sin \beta \sin \phi + g \cos \beta (\cos \phi \cos \theta \cos \alpha) \\ -\cos \alpha \sin \theta - \frac{\bar{q}S}{m} (C_{D_0} + C_{D_2} \alpha^2) \\ R^{-1}S^\times(\omega)R\omega + \bar{q}SR^{-1}f(x_2, x_3) \end{bmatrix} \quad (30)$$

and $B_3 : R^2 \times (0, \infty) \rightarrow R^{4 \times 4}$ is the input matrix function for the inner actuated dynamics and is given by

$$B_3(x_2, \bar{q}) = \begin{bmatrix} 0_{1 \times 3} & \frac{T_{\max} \cos(\alpha + \delta) \cos \beta}{m} \\ & \bar{q}SR^{-1}B_\omega \end{bmatrix} \quad (31)$$

3 Controller Design

The model is given as $x = [x_1^T \ x_2^T \ x_3^T]^T = [\phi \ \theta \ \psi \ \alpha \ \beta \ Vp \ q \ r]^T$

$$\dot{x} = \tilde{A}(x) + \tilde{B}(x)u + \Delta(x) \quad (32)$$

where $\Delta(x)$ is an uncertainty due to failure, and

$$\tilde{A}(x) = \begin{bmatrix} A_1(x_1) \\ A_2(x_1, x_2, x_3) \\ A_3(x_1, x_2, x_3) \end{bmatrix}, \tilde{B}(x) = \begin{bmatrix} 0 \\ B_2(x_2, \bar{q}) \\ B_3(x_2, \bar{q}) \end{bmatrix} \quad (33)$$

The DI controller computes the actuator commands from the desired acceleration input \dot{x}_d as

$$u_c = \tilde{B}^*[\dot{x}_d - \tilde{A}(x_d)] \quad (34)$$

where

$$\tilde{B}^*(x) = \tilde{B}^T(x)[\tilde{B}(x)\tilde{B}^T(x) + \tau(t)I_{9 \times 9}]^{-1} \quad (35)$$

and $\tau(t) : [0, \infty) \rightarrow R$ satisfies Eq. (35).

$$\dot{x}_d = \dot{x}_m + K_p(x_m - x) + K_i \int_0^t (x_m - x) d\tau - u_{ad} \quad (36)$$

The adaptive element u_{ad} is designed to cancel out the uncertainty term $\Delta(x)$ and is linearly parameterized as

$$u_{ad} = E^T \Phi(x) \quad (37)$$

4 Stability Analysis

Assuming no actuator rate limiting and saturation, then the control surface deflection commands will faithfully follow the actuator commands. The tracking error equation is then formed by

$$\dot{e} = A \cdot e + B[\tilde{E}^T \Phi(x) - \varepsilon(x)] \quad (38)$$

where $\tilde{E} = E - E^*$ is a parameter error, $\varepsilon(x)$ is an approximation error,

$$e = \left[\int_0^t (x_m - x) d\tau \quad (x_m - x) \right]^T$$

is a tracking error, and

$$A = \begin{bmatrix} 0 & I \\ -K_i & -K_p \end{bmatrix}, B = \begin{bmatrix} 0 \\ I \end{bmatrix} \quad (39)$$

The weight update law based on the OCM is given as

$$\dot{E}^T = -\Gamma \Phi(x) [e^T P - v \Phi^T(x) E B^T P A^{-1}] B \quad (40)$$

where $v > 0$ is a modification parameter, $\Gamma = \Gamma^T > 0$ is an adaptive gain matrix, and $P = P^T > 0$ solves

$$PA + A^T P = -Q \quad (41)$$

Let $Q = 2I$, then the solution of P is given as

$$P = \begin{bmatrix} K_i^{-1} K_p + K_p^{-1} (K_i + I) & K_i^{-1} \\ K_i^{-1} & K_p^{-1} (K_i^{-1} + I) \end{bmatrix} \quad (42)$$

Then, the term $B^T P A^{-1} B$ is evaluated as

$$B^T P A^{-1} B = -(K_i^{-1})^2 < 0 \quad (43)$$

The alternative form of the OCM weight update law is given as

$$\dot{E}^T = -\Gamma\Phi(x)[e^T PB + v\Phi^T(x)E(K_i^{-1})^2] \quad (44)$$

Stability of the OCM adaptive law can be established by a Lyapunov proof as follows.

Choose a Lyapunov function

$$V(e, \tilde{E}) = e^T P e + \text{trace}(\tilde{E}^T \Gamma^{-1} \tilde{E}) \quad (45)$$

Thus, evaluating $\dot{V}(e, \tilde{E})$ yields

$$\begin{aligned} \dot{V}(e, \tilde{E}) = & -e^T Q e - 2e^T P B \varepsilon - v\Phi^T(x)\tilde{E}B^T A^{-T} Q A^{-1} B \tilde{E}^T \Phi(x) \\ & + 2v\Phi^T(x)E^* B^T P A^{-1} B \tilde{E}^T \Phi(x) \end{aligned} \quad (46)$$

$\dot{V}(e, \tilde{E})$ is then bounded by

$$\begin{aligned} \dot{V}(e, \tilde{E}) \leq & -\|e\|[\lambda_{\min}(Q)\|e\| - 2\|PB\|\varepsilon_0] - v\|\Phi^T(x)\|^2\|\tilde{E}\|^2 \\ & [\lambda_{\min}(B^T A^{-T} Q A^{-1} B)\|\tilde{E}\| - 2\|B^T P A^{-1} B\|E_0^*] \end{aligned} \quad (47)$$

where $\varepsilon_0 = \sup \| \varepsilon(x) \|$ and $E_0^* = \max \| E^* \|$. Let

$$\begin{aligned} c_1 & \doteq \lambda_{\min}(Q), \quad c_2 \doteq \frac{\|PB\|\varepsilon_0}{\lambda_{\min}(Q)}, \quad c_3 \doteq \lambda_{\min}(B^T A^{-T} Q A^{-1} B)\|\Phi(x)\|^2, \\ c_4 & \doteq \frac{\|B^T P A^{-1} B\|E_0^*}{\lambda_{\min}(B^T A^{-T} Q A^{-1} B)}. \end{aligned} \quad (48)$$

Then, $\dot{V}(e, \tilde{E}) \leq 0$ implies either

$$\|e\| \geq r = c_2 + \sqrt{c_2^2 + \frac{vc_3c_4^2}{c_1}}, \quad \text{or} \quad \|\tilde{E}\| \geq \alpha = c_4 + \sqrt{c_4^2 + \frac{vc_1c_2^2}{vc_3}} \quad (49)$$

There exists a maximum value v_{\max} such that $0 \leq v < v_{\max}$ for which

$$\begin{aligned} \varphi(\|x\|, \|x_m\|, Q, v, \varepsilon_0, E_0) = & -c_1\|x\|^2 + 2(c_1c_2 + c_5\|x_m\|)\|x\| + 2c_1c_2\|x_m\| \\ & - c_1\|x_m\|^2 + vc_3(\|\Phi(x)\|)c_4^2 \leq 0 \end{aligned} \quad (50)$$

where $c_5 \doteq \lambda_{\max}(Q)$.

Then, $\|\Phi(x)\|$ has an upper bound where

$$\|\Phi(x)\| \leq \|\Phi(\varphi^{-1}(\|x_m\|_{\infty}, Q, v, \varepsilon_0, E_0))\| = \Phi_0 \quad (51)$$

which implies $c_3 = \lambda_{\min}(B^T A^{-T} Q A^{-1} B)\Phi_0^2$. Let

$$B_{\delta} = \{(e, \tilde{E}) : c_1(\|e\| - c_2)^2 + vc_3(\|\tilde{E}\| - c_4)^2 \leq c_1c_2^2 + vc_3c_4^2\} \quad (52)$$

Then, $\dot{V}(e, \tilde{E}) > 0$ inside of B_{δ} , but $\dot{V}(e, \tilde{E}) \leq 0$ outside of B_{δ} . Therefore, $\dot{V}(e, \tilde{E})$ is a decreasing function outside of B_{δ} . All trajectories $(e(0), \tilde{E}(0))$ will be ultimately bounded after some time $t > T$ with the following ultimate bounds:

$$\|e\| \leq \rho = \sqrt{\frac{\lambda_{\max}(P)r^2 + \lambda_{\max}(\Gamma^{-1})\alpha^2}{\lambda_{\min}(P)}} \quad (53)$$

$$\|\tilde{E}\| \leq \eta = \sqrt{\frac{\lambda_{\max}(P)r^2 + \lambda_{\max}(\Gamma^{-1})\alpha^2}{\lambda_{\min}(\Gamma^{-1})}} \quad (54)$$

for any $0 \leq v < v_{\max}$, such that $\varphi(\|x\|, \|x_m\|, Q, v, \varepsilon_0, E_0) \leq 0$.

The role of the modification parameter v is important. If tracking performance is more desired in a control design than robust stability, then a small value of v should be selected. In the limit when $v = 0$, the standard MRAC is recovered and asymptotic tracking performance is achieved but at the expense of robust stability. The lack of robust stability of the standard MRAC is well understood. On the other hand, if robust stability is a priority in a design, then a larger value of v should be chosen. As with any control design, tracking performance and robust stability are often considered as two competing design requirements. By judiciously selecting the modification parameter v , the OCM adaptive law can be designed to achieve a specified level of tracking performance while maintaining sufficient robust stability.

5 Conclusion

The adaptive OCM/NN-based flight control design technique considered in this paper as a control assistant to the DI controller to perform rate stabilization. It is premature at this stage to conclude that the OCM/NN controller is the best control augmenter among its adaptive design counterparts (e.g., MRAC as design add-on to PID or SDRE for inner loop compensation, [13]); nevertheless, its inherent ability to interact with the UAV's primary controller (DI controller in this case) in order to maintain the vehicle stability while the baseline fails to stabilize deserves more attention in the control community.

References

1. Ren W, Beard R (2004) Trajectory tracking for unmanned air vehicles with velocity and heading rate constraints. AIAA Aerospace Conference. IEEE Trans Control Syst Technol 12(5):706–716
2. Valavanis K, Oh P, Piegel L (eds) (2009) Unmanned aircraft systems. Springer, London
3. Wegener S, Sullivan D, Frank J, Enomoto F (2004) UAV Autonomous operations for airborne science missions. In: AIAA 3rd Unmanned Unlimited Technical Conference, Workshop and Exhibit: 2004–6416
4. Gruszka A, Malisoff M (eds) (2012) Bounded tracking controllers and robustness analysis for UAVs. IEEE Tran Autom Control: 19–21

5. Shima T, Rasmussen S (eds) (2009) UAV cooperative decision and control: challenges and practical approaches. SIAM, Philadelphia
6. Ailon A (2009) Trajectory tracking for UAVs with bounded inputs and some related applications. IFAC Symposium on Robust Control Design, Haifa, Israel: 355–360
7. Tony (2009) Six-DOF trajectory tracking for payload directed flight using trajectory linearization control. In: AIAA Aerospace Conference, Washington: 1897
8. AICHiddabi SA, McClamroch NH (2002) Aggressive longitudinal aircraft trajectory tracking using nonlinear control. *J Guidance Control Dyn* 25(1): 26–32
9. Fujimori A, Kurozumi M, Nikiforuk PN, Gupta MM (2000) Flight control design of an automatic landing flight experiment vehicle. *J Guidance Control Dyn* 23(2):373–376
10. Sieberling S, Chu QP, Mulder JA (2010) Robust flight control using incremental nonlinear dynamic inversion and angular acceleration prediction. *J Guidance Control Dyn* 33(6):1732–1742
11. Hameduddin I, AH (2012) Bajodah nonlinear generalised dynamic inversion for aircraft manoeuvring control. *Int J Control* : 1–14
12. Lam Quang M, Nguyen Nhan T, Oppenheimer Michael W (2012) Intelligent adaptive flight control using optimal control modification and neural network as control augmentation layer and robustness enhancer. In: AIAA Aerospace Conference, California: 19–21
13. Schierman John D, Ward David G, Hull Jason R, Gandhi Neha (2004) Integrated adaptive guidance and control for re-entry vehicles with flight-test results. *J Guidance Control Dyn* 27(6):975–988
14. Ochi S, Takano H, Baba Y (2002) Flight trajectory tracking system applied to inverse control for aerobatic maneuvers. *Inverse problems in engineering mechanics*. Elsevier Science Ltd, Amsterdam: 337–344
15. Beard R, McLain T, Goodrich M, Anderson E (2002) Coordinated target assignment and intercept for unmanned air vehicles. *IEEE Trans. Robot Autom* 18(6):911–922
16. Gu G, Chandler P, Schumacher C, Sparks A, Pachter M (2006) Optimal cooperative sensing using a team of UAVs. *IEEE Trans. Aerosp Electro Syst* 42(4): 1446–1458
17. Valavanis K, Oh P, Piegel L (eds) (2009) Unmanned aircraft systems. Springer, London
18. Jiang Z-P, Lefeber E, Nijmeijer H (2001) Saturated stabilization and tracking control of a nonholonomic mobile robot. *Syst. Control Lett.* 42(5): 327–332
19. Park S, Deyst J, Howz JP (2004) A new nonlinear guidance logic for trajectory tracking. AIAA: 2004–1430
20. Lane SH, Stengel RF (1988) Flight control design using non-linear inverse dynamics. *Automatica* 24:471–483
21. Yoon H, Agrawal BN (2009) Adaptive control of uncertain Hamiltonian multi-input multi-output systems: with application to spacecraft control. *IEEE Trans Control Syst Technol* 17(4):900–906

Evolutionary Game Analysis on Enterprise's Knowledge-Sharing in the Cooperative Networks

Shengliang Zong, Zhen Cai and Mingyue Qi

Abstract Knowledge resource is an important strategic resource of the enterprise, and the advantage of knowledge resources determines the enterprise's competitiveness. Based on evolutionary game theory, the paper studies the evolutionary direction of enterprise's knowledge-sharing in the cooperative networks and analyzes the influencing factors. Our analysis shows that the evolutionary direction of the system has a connection with the bilateral game payoff matrix and that the system's original state also affects the result under non-contract condition. The probability of knowledge-sharing of both sides has a positive correlation with excess result and a negative correction with cost of knowledge-sharing and betrayal income, and allocation factor of excess result is the key influence factor. In the case of an alliance contract, the evolutionary direction of the system is knowledge-sharing of both sides because of the existence of punishment mechanism.

Keywords Innovation networks · knowledge-sharing · Evolutionary game

S. Zong (✉) · Z. Cai
School of Management, Lanzhou University, Lanzhou 730000, China
e-mail: zongshl03@gmail.com

Z. Cai
e-mail: caizh07@lzu.cn

M. Qi
Ecole de Management, Institute Mines-Telecom & Management Sudparis Evry,
Evry 91000, France
e-mail: qimingyue1986@hotmail.com

1 Introduction

Knowledge has become the important source of the construction of the competitive advantage in knowledge economic era. The enterprises try their best to find the cooperative chance in order to obtain the resource and knowledge to enhance their cooperative ability. Thus, the cooperative innovation network has become the main path for the enterprises to obtain the study resource.

Schneider [1] studied the development of no-tillage by combining concepts of co-creation of knowledge and actor-network theory. The reconstruction of the process of no-tillage development in Switzerland has made it possible to show that no-tillage development may be regarded as a dynamic process of co-creation of innovation. Spencer [2] explored the relationship between firms' strategies sharing knowledge with their innovation system and their innovative performance. The empirical analysis showed that many firms designed strategies to share technological knowledge with their competitors, and those firms sharing knowledge with their innovation system earned higher innovative performance than firms did not do so. In addition, firms interacting with their global innovation system earned higher innovative performance than firms interacting only with their national innovation system. Wang [3] investigated the quantitative relationship between knowledge-sharing, innovation, and performance. He developed a research model and posited that knowledge-sharing not only has positive relationship with performance directly but also affects innovation which in turn contributes to firm performance. This model is empirically tested using data collected from 89 high-technology firms in Jiangsu Province of China. It found that both explicit and tacit knowledge-sharing practices facilitated innovation and performance. Explicit knowledge-sharing has more significant effects on innovation speed and financial performance, while tacit knowledge-sharing has more significant effects on innovation quality and operational performance.

Camelo Ordaz [4] studied the relationship between knowledge-sharing and innovation. They pursued two aims: firstly, to identify knowledge-sharing enablers, and secondly, to analyze the effect of knowledge-sharing processes on innovation performance. Hsiu-Fen [5] proposed to examine the influence of individual factors (enjoyment in helping others and knowledge self-efficacy), organizational factors (top management support and organizational rewards), and technology factors (information and communication technology use) on knowledge-sharing processes and investigate that whether more leads to superior firm innovation capability.

In this paper, we studied the knowledge-sharing problem of enterprises in innovation networks using evolution game theory and method. Our advantage is that we treat the innovation networks as a progressive evolution system of 'learning' and emphasize its dynamic and macro. Firstly, we constructed the evolution game model of knowledge-sharing in innovation networks without contract. Furthermore, we analyzed the factors that influence the stability of the system. Then, we constructed the evolution game model with contract. At last, we got the useful conclusions through analyses of the model.

2 Variables Defined and Construct of Game Model

In order to study, we divided the enterprises into big enterprise and small enterprise according to their scale. The game strategies of them are all sharing or non-sharing. The symbol and the variables are defined as following:

- L Core enterprise.
- S Non-core enterprise.
- π_c The normal revenue of core enterprise with knowledge non-sharing.
- π_n The normal revenue of non-core enterprise with knowledge non-sharing.
- $\Delta\pi$ Excess revenue with both core enterprise and non-core enterprise knowledge-sharing.
- w Allocation factor of core enterprise in excess revenue with knowledge-sharing.
- c_c The cost of core enterprise for knowledge-sharing.
- c_n The cost of non-core enterprise for knowledge-sharing.
- k Betrayal cost that do not take knowledge-sharing.

Table 1 shows the game payoff matrix of core enterprise and non-core enterprise of innovation networks.

Assume that α denote the ratio of knowledge-sharing that core enterprise has chosen. Then, $1 - \alpha$ denote the ratio of knowledge non-sharing. Meanwhile, assume that β denote the ratio of knowledge-sharing that non-core enterprise has chosen, and $1 - \beta$ denote the ratio of knowledge non-sharing. Thus, we have the fitness $u_{cs}, u_{c\bar{s}}$, and the average fitness \bar{u}_c of core enterprise for knowledge-sharing and non-sharing, respectively.

$$u_{cs} = \beta(\pi_c + \Delta\pi w - c_c) + (1 - \beta)(\pi_c - c_c) \tag{1}$$

$$u_{c\bar{s}} = \beta(\pi_c + \Delta\pi w - c_c) + (1 - \beta)(\pi_c - c_c) \tag{2}$$

$$\bar{u}_c = \alpha u_{cs} + (1 - \alpha)u_{c\bar{s}} \tag{3}$$

Thus, the replicated dynamic equation of core enterprise that chosen knowledge-sharing is

$$\frac{d\alpha}{dt} = \alpha(u_{cs} - \bar{u}_c) = \alpha(1 - \alpha)[\beta(\Delta\pi w - k) - c_c] \tag{4}$$

Table 1 Game payoff matrix of core enterprise and non-core enterprise

		Non-core enterprise S	
		Sharing	Non-sharing
Core enterprise L	Sharing	$\pi_c + \Delta\pi w - c_c,$ $\pi_n + \Delta\pi(1 - w) - c_n$	$\pi_c - c_c, \pi_n + k$
	Non-sharing	$\pi_c + k, \pi_n - c_n$	π_c, π_n

Similarly, the replicated dynamic equation of non-core enterprise that chose knowledge-sharing is

$$\frac{d\beta}{dt} = \beta(1 - \beta)\{\alpha[\Delta\pi(1 - w) - k] - c_n\} \tag{5}$$

Differential Eqs. (4) and (5) described the population dynamics of this game system. When $\frac{d\alpha}{dt} = 0$, we have $\alpha = 0$, $\alpha = 1$ or $\beta = \frac{c_c}{\Delta\pi w - k}$. Similarly, we have $\beta = 0$, $\beta = 1$ or $\alpha = \frac{c_n}{\Delta\pi(1-w) - k}$, while $\frac{d\beta}{dt} = 0$.

Thus, we can get five equilibrium points $O(0, 0)$, $A(1, 0)$, $B(0, 1)$, $C(1, 1)$ and $D\left(\frac{c_n}{\Delta\pi(1-w) - k}, \frac{c_c}{\Delta\pi w - k}\right)$, above the plane $S = \{(\alpha, \beta); 0 \leq \alpha, \beta \leq 1\}$. The stability of these equilibrium points could be obtained through the analysis of the local stability of the Jacobian matrix of this system, according to the method that Fridman [6] proposed. Then, the Jacobian matrix of this system is

$$J = \begin{pmatrix} (1 - 2\alpha)[\beta(\Delta\pi w - k) - c_c] & \alpha(1 - \alpha)(\Delta\pi w - k) \\ \beta(1 - \beta)[\Delta\pi(1 - w) - k] & (1 - 2\beta)\{\alpha[\Delta\pi(1 - w) - k] - c_n\} \end{pmatrix}.$$

We have $J = \begin{pmatrix} -c_c & 0 \\ 0 & -c_n \end{pmatrix}$, $\det J = (-c_c)(-c_n) > 0$, and $trJ = (-c_c) + (-c_n) < 0$ at the equilibrium point $O(0, 0)$. Thus, equilibrium point $O(0, 0)$ is the evolution stable state. Table 2 shows the result of the local stability for all these equilibrium points. We can obviously find that $O(0, 0)$ and $C(1, 1)$ are evolution stable points, and $A(1, 0)$ and $B(0, 1)$ are astable points, and $D\left(\frac{c_n}{\Delta\pi(1-w) - k}, \frac{c_c}{\Delta\pi w - k}\right)$ is saddle point.

Furthermore, we can get the dynamic evolution phase diagram of the enterprise’s knowledge-sharing behavior in innovation networks.

Figure 1 shows that the polyline consists of astable points $A(1, 0)$, $B(0, 1)$ and saddle point $D\left(\frac{c_n}{\Delta\pi(1-w) - k}, \frac{c_c}{\Delta\pi w - k}\right)$, which divided the system into two parts. The system will astringe to $O(0, 0)$, when the initial state is in the region of F , which means that both of them do not take knowledge-sharing action. Similarly, the system will astringe to $C(1, 1)$, when the initial state is in the region of E , which means that they will take knowledge-sharing act. Sharing and non-sharing may coexist for a long time because the evolution of the system is a long process.

Table 2 The result of the local stability analysis for the system

Equilibrium points	detJ	trJ	Result
$O(0, 0)$	+	−	ESS
$A(1, 0)$	+	+	Unstable
$B(0, 1)$	+	+	Unstable
$C(1, 1)$	+	−	ESS
$D\left(\frac{c_n}{\Delta\pi(1-w) - k}, \frac{c_c}{\Delta\pi w - k}\right)$	−	0	Saddle point

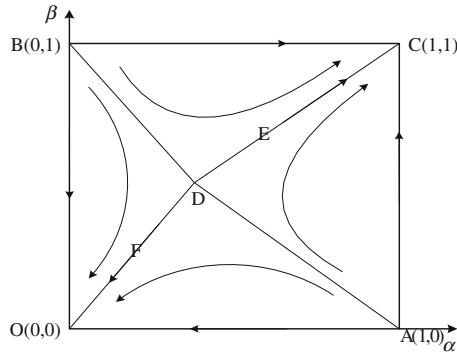


Fig. 1 The dynamic evolution phase diagram of the enterprise's knowledge-sharing behavior in innovation networks

3 Model Analyze

We can get the following results from the above analysis. Firstly, the equilibrium result of the system may be sharing or non-sharing, and it depends on the payoff matrix. Secondly, the convergence point of the system is decided by the initial state. Therefore, the initial value of some parameters and its change will affect the evolution path, which makes the system convergence in different directions. In order to study, we take F for example.

$$S_F = \frac{1}{2} \left(\frac{c_n}{\Delta\pi(1-w) - k} + \frac{c_c}{\Delta\pi w - k} \right) \tag{6}$$

It is obvious that $\Delta\pi, k, w, c_c$ and c_n will affect the value of S_F .

- (1) Excess revenue $\Delta\pi$. We can get that S_F is decreasing with $\Delta\pi$ increasing, which means that the probability of the system convergence to $C(1,1)$ is increasing. The knowledge resource that the enterprise owned is highly complementary. We can obtain the maximum excess revenue by realizing the knowledge-sharing, which makes the relationship among enterprises more stability.
- (2) The cost c_c and c_n . We can get that S_F is increasing with c_c and c_n increasing, which means that the probability of the system convergence to $O(0,0)$ is increasing. It shows that the enterprise will not take knowledge-sharing act.
- (3) Betray revenue k . We can get that S_F is increasing with k increasing, which means that the probability of the system convergence to $O(0,0)$ is increasing. It shows that the enterprise will not take knowledge-sharing act. The enterprise will continue to take non-sharing act if the betray revenue is big enough when it takes non-sharing act. While the sharing enterprise that who will pay a cost cannot obtain the share revenue, it will take retaliatory strategy.

(4) Allocation factor w . When other parameters are fixed, we can get the first derivative about w , and we have $\frac{\partial S_F}{\partial w} = \frac{1}{2} \Delta\pi \left(\frac{c_n}{[\Delta\pi(1-w)-k]^2} - \frac{c_c}{(\Delta\pi w-k)^2} \right)$. When $\frac{c_n}{[\Delta\pi(1-w)-k]^2} = \frac{c_c}{(\Delta\pi w-k)^2}$, we will get the minimum of S_F , and the probability of the system convergence to $O(0, 0)$ is increasing. It shows that the enterprise will not take knowledge-sharing act.

4 Evolution Analyses under Contract

We know that the evolution path of the system is uncertain without contract. Trust mechanism and contract mechanism are two main factors that affect the cooperative relations. Contract mechanism will pay an important part to maintain the good market order while the trust mechanism fails. The trust mechanism has been proved to be useful for long cooperative, and it can reduce the cooperative cost and avoid the opportunism.

Enterprises will sign a contract before the cooperation. The non-sharing parts will be punished and the sharing parts will be rewarded. Enterprises will pay a cost for signing the contract if all of them do not take sharing act. Table 3 shows the game revenue matrix of enterprises under contract.

Thus, we can get the replicated dynamic equation of core enterprise and non-core enterprise, respectively,

$$\frac{d\alpha}{dt} = \alpha(1 - \alpha)(\Delta\pi\beta w + u + c_c^c - c_c) \tag{7}$$

$$\frac{d\alpha}{dt} = \alpha(1 - \alpha)(\Delta\pi\beta w + u + c_c^c - c_c) \tag{8}$$

Differential Eqs. (7) and (8) described the population dynamics of this game system. Thus, we can get four equilibrium points $O(0, 0), A(1, 0), B(0, 1), C(1, 1)$ above the plane $S = \{(\alpha, \beta); 0 \leq \alpha, \beta \leq 1\}$. Then, we get $O(0, 0)$ that is astable point, and $A(1, 0)$ and $B(0, 1)$ are saddle points, and $C(1, 1)$ is evolution stable point.

Table 3 The game revenue matrix under contract

		Non-core enterprise S	
		Sharing	Non-sharing
Core enterprise L	Sharing	$\pi_c + \Delta\pi w - c_c$ $\pi_n + \Delta\pi(1-w) - c_n$	$\pi_c - c_c + u,$ $\pi_n - u - c_n^c$
	Non-sharing	$\pi_c - u - c_c^c, \pi_n - c_n + u$	$\pi_c - c_c^c, \pi_n - c_n^c$

u : The punishment that non-sharing enterprise pays for sharing enterprise

c_c^c : The cost that the core enterprise paid for signing the contract

c_n^c : The cost that the non-core enterprise paid for signing the contract

Figure 2 shows that the evolution result of the system is to take sharing strategy under contract condition among enterprises in innovation networks. If the initial state is non-sharing strategies, anyone of them will take act to change the state in order to obtain the excess revenue. If the initial state is one part sharing and another part non-sharing, the sharing part will continue to take sharing strategy and the non-sharing part will change their non-sharing strategy in order to obtain the normal revenue. If the initial state is sharing strategy, both of them will continue to take their strategies because they can obtain more revenue. At last, both of the game parts will keep state on sharing strategies by learning and adjusting.

5 Numeral Examples

The enterprises can get the normal revenue in the cooperative innovation networks. The core enterprise and non-core enterprise can make the normal revenue π_c and π_n , respectively. We find that they do not have any effect on the result of the evolution. Thus, we do not consider them in the numeral examples.

(1) knowledge-sharing without contract

We suppose there is a cooperative network that contains core enterprises and non-core enterprises. The core enterprise who designs the contract plays an important part in this cooperative network. The excess revenue for knowledge-sharing is 1,800, and the allocation factor of core enterprise in excess revenue with knowledge-sharing is 0.60. The cost of core enterprise and non-core enterprise for knowledge-sharing is 100 and 60, respectively. The betrayal cost that does not take knowledge-sharing while another one takes knowledge-sharing is 300.

We can get the result that the probability of the knowledge-sharing between core enterprise and non-core enterprise is 0.864. The net revenues of the core

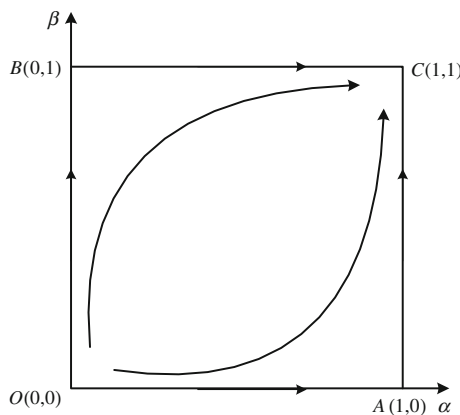


Fig. 2 The dynamic evolution phase diagram of the enterprise's knowledge-sharing behavior under contract

enterprise and non-core enterprise are 980 and 660, respectively. The allocation factor is decided through the negotiations between the enterprises. The allocation factor that was given is not the optimal value so that the probability is not the biggest. We can acquire the optimal allocation factor that is 0.5423, and the biggest probability of the knowledge-sharing between the enterprises is 0.869. The net revenues of the core enterprise and non-core enterprise are 876.14 and 763.86, respectively. It indicates that the core enterprise should transfer a part of its revenue to the non-core enterprise in order to obtain the biggest probability for knowledge-sharing.

(2) knowledge-sharing with contract

We suppose that the enterprises signed a contract at the beginning of the cooperation. The costs that the core enterprise and non-core enterprise paid for signing the contract are 40 and 20, respectively. The punishment that non-knowledge-sharing enterprise paid for knowledge-sharing enterprise is 150. $c_c^c = 40$, $c_n^c = 20$, $c_c = 100$, $c_n = 60$, and $u = 150$. We make the parameter value into the differential Eqs. (7) and (8). We can find them meet the conditions $u > c_c - c_c^c$ and $u > c_u - c_u^c$. Therefore, the evolution result of the system is that both of the core enterprise and non-core enterprise will take knowledge-sharing strategy under contract among enterprises in innovation networks.

6 Conclusions

In this paper, we studied the evolutionary paths of enterprise knowledge-sharing in innovation networks by using evolutionary game method. The game direction of the system is related to the payoff matrix and is affected by the initial state of the system without contracts. At the same time, excess revenue, sharing cost, betray revenue, and allocation factor are important affect factors. The system may converge to sharing strategies or non-sharing strategies for a long time. The system would converge to sharing strategies under contracts.

Acknowledgments This work is supported by the Fundamental Research Funds for the Central Universities (Project NO. 11LZUJBWZY098).

References

1. Schneider F, Steiger D, Ledermann T, Fry P, Rist S (2012) No-tillage farming: co-creation of innovation through network building. *Land Degrad Dev* 23(3):242–255
2. Spencer JW (2003) Firms' knowledge-sharing strategies in the global innovation system: empirical evidence from the flat panel display industry. *Strategic Manage J* 24:217–233
3. Wang ZN, Wang NX (2012) knowledge-sharing, innovation and firm performance. *Expert Syst Appl* 39:8899–8908

4. Ordaz CC, Cruz JG, Ginel ES (2010) knowledge-sharing: enablers and its influence on innovation. *Cuadernos De Economia Y Direccion De La Empresa*, pp 113–150
5. Lin HF (2007) knowledge-sharing and firm innovation capability: an empirical study. *Int J Manpower* 28:315–332
6. Friedman D (1991) Evolutionary games in economics. *Econometrician* 59:637–639

An Ant Colony Algorithm for Two-Sided Assembly Line Balancing Problem Type-II

Zeqiang Zhang, Junyi Hu and Wenming Cheng

Abstract Two-sided assembly lines are widely used in the assembly of large-sized products, such as automobiles, buses, or trucks. Two-sided assembly line problem is more difficult than single-sided assembly line problem as the operation directions constrain of tasks and the requirement of parallel work in the task distribution procedure. In this paper, the mathematical model of two-sided assembly line balancing problem type-II (TALBP-II) is given first. And then, an improved ant colony algorithm was constructed to solve TALBP-II. In the algorithm, a hybrid ant-based search rule and a heuristic task distribution rule were used in order to establish a feasible solution, global pheromone trail update, and the optimum solution search strategy are also considered. The feasibility of this algorithm was indicated by a case of a loader final assembly line.

Keywords Two-sided assembly lines · Balancing · Ant colony algorithm

1 Introduction

Considering the product's size and the complexity of the assembly line, two-sided assembly lines are widely used in assembly of large-sized products such as automobiles, buses, or trucks. It is very important for manufacturers to design an efficient two-sided assembly line. Two-sided assembly line balancing problem (TALBP) can be divided into two versions: TALBP-I consists of assigning tasks to

Z. Zhang (✉) · W. Cheng
School of Mechanical Engineering, Southwest Jiaotong University,
610031 Chengdu, China
e-mail: zhagnzeqiang@gmail.com

J. Hu
CSR Qishuyan Institute Co., Ltd, 213011 Changzhou, China

work stations, such that the number of open position of the two-sided assembly line is minimized for a given production rate, and TALBP-II is to minimum cycle time for a given number of open position in the assembly line. In actual manufacturing enterprises, the optimization goals often refer to the second category in order to reduce the production cycle time.

TALBP are received widely attention in the past few years for the importance of two-sided assembly lines. Bartholdi [1] first proposed two-sided assembly line balancing problem, and a first-fit rule-based heuristic algorithm is given for TALBP-I. Kim et al. [2] and Lee et al. [3] used genetic algorithm for solving TALBP-I and join the group distribution strategy to optimize the correlation indicators between tasks. Simaria and Vilarinho [4] used ant colony optimization algorithm, named 2-ANTBAL, for solving TALBP-I. Baykasoglu and Dereli [5] first introduced the concept of regional constraints in TALBP-I and then proposed an ant colony algorithm to solve the problem. Wu and Hu proposed a branch-and-bound algorithm for TALBP-I [6].

Compare to TALBP-I, TALBP-II is less studied [7]. Ozcan [8] presented a Petri net-based heuristic to solve simple assembly line balancing problem type-II. To solve TALBP-II, Kim et al. [9] presented a mathematical model and a genetic algorithm, which adopted the strategy of localized evolution and steady-state reproduction to promote population diversity and search efficiency. In this paper, a new ant colony algorithm is proposed to solve the two-sided assembly line balance problem as the common problems of manufacturing enterprises, respectively, using the ant colony comprehensive search rules and task assignment rules to select and assign tasks. Considering the special features of TALBP-II, an ant colony the overall search process is introduced in this paper.

This paper is organized as follows: The TALBP-II is discussed in Sect. 2, and mathematical model is presented. In Sect. 3, the modified ant colony algorithm for TALBP-II is explained in detail. An example problem of a loader final assembly line is solved using the proposed method in Sect. 4. Finally, some conclusions are presented in Sect. 5.

2 Problems Statement and Formulation

As shown in Fig. 1, in two-sided assembly line, both sides of the station can operate different tasks in parallel. Stations 1 and 2 can be described as a pair of

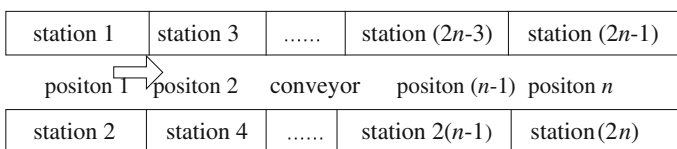


Fig. 1 Two-sided assembly line

stations (mated-station), the station 1 can be called an accompanied station of the station 2 (a companion), and the remaining work stations have the similar relationship. Tasks can be operated simultaneously when they do not have priority constraints.

The mathematical model of two-sided assembly line balancing problem can be described as below. First, list the relevant variables: I —task set, $I = \{1, 2, \dots, N\}$; I_D —to split the task set by the task operating position, $D = L(\text{left}), R(\text{right}),$ or $E(\text{both sides}), I_E$ refers to the tasks can be assigned to both sides of the position, and so on; $I_j' I_j'$ —sets of tasks assigned to station j and j' ; $T_j' R_j'$ —the total operating time and total delay time of station j ; T_j', R_j' —total operating time and total delay time of station j' ; $P(i)$ —set of all predecessors of task i ; S —set of all the task i whose $P(i)$ is empty, all this tasks can be used for distribution into the station because their pre-order tasks are already located; S' —set of tasks meet the cycle time constraints from S , the tasks can be assigned to the station directly; W —set of unassigned tasks; n —the number of opened position; N_S —opened work stations; x_{ipk} —the value is 1, if task i is assigned to position p , and the operating position is $k, k \in \{L, R\}$; 0, otherwise.

Constraints:

$$\text{If } x_{ipk} = 1, \text{ then } x_{hgk} = 0, h \in P(i), g, p = 1, 2, 3, \dots, n, \text{ and } g > p \quad (1)$$

$$\sum_{i \in I} \sum_{p=1}^n x_{ipk} = 1, k \in \{L, R\} \quad (2)$$

$$T_j = \sum_{i \in I} x_{ipk} \cdot t_i, \text{ and } k = L, j = 2p - 1 \quad (3)$$

$$T_j' = \sum_{i \in I} x_{ipk} \cdot t_i, \text{ and } k = R, j' = 2p \quad (4)$$

$$\text{In any position } p, T_j + R_j \leq C, \text{ and } T_j' + R_j' \leq C \quad (5)$$

Objective:

$$\text{Minimize } C \quad (6)$$

$$\max \text{LE} = \left(\sum_{i \in I} t_i \right) / (N_S C) \quad (7)$$

$$\min \text{SI} = \sqrt{\frac{\sum_{i=1}^{N_S} (T_{\max} - T_i)^2}{N_S}} \quad (8)$$

Lower limit formulation of cycle time:

$$C_{\min} = \max\{|T_{\text{SUM}}/2n|, |T_{\text{SUML}}/n|, |T_{\text{SUMR}}/n|, t_{\max}\} \quad (9)$$

Constraint (1) is the assignment constraint which ensures that each task is assigned to exactly one station and all precedence relations between tasks are satisfied. Constraint (2) ensures the task must be assigned to a position p , and the allocated operating position only can be left or right. Constraints (3) and (4), respectively, refers to total operating time of station j and j' . Constraint (5) refers to the operating time of any mated-stations in any position p must satisfy the cycle time constraints. Objective function (6) as the main objective function this paper focused on TALBP-II, that is to minimize the cycle time when the number of positions opened in the assembly lines is determined. Functions (7) and (8) are secondary evaluation objective, to maximum line efficiency LE and minimum smoothness index SI can balance the stations' load. When $LE = 1$ and $SI = 0$, every position has the same load. Formulation (9) is lower limit formulation of cycle time for TALBP-II; T_{SUM} is the sum of task time in the set I ; T_{SUML} , T_{SUMR} , and T_{SUME} were the sum operating time of all tasks in I_L , I_R , and I_E , respectively; t_{max} is the task which has the longest operating time.

3 The Ant Colony Algorithm for TALBP-II

3.1 Ant Colony Comprehensive Search Rules

Selected tasks from the set S which have the earliest start time. These tasks constitute FT (the tasks in FT have the same earliest start time); if FT has only one task, select this task by task assignment rules. If FT has more than one tasks, then use the hybrid search mechanism to select the task, this mechanism borrow ideas called improved pheromone rule from the literature [10]:

$$i = \begin{cases} I_1 : p_{ij} = \frac{\left(\sum_{k=1}^j \tau_{ih}\right)^\alpha (\eta_i)^\beta}{\sum_{s \in FT} \left(\sum_{h=1}^j \tau_{sh}\right)^\alpha (\eta_s)^\beta} & 0 \leq r \leq r_1 \\ I_2 : \text{random selection of } i \in FT & r_1 < r \leq 1 \end{cases} \quad (10)$$

r —random number between (0, 1); r_1 —the user-defined parameter to meet $0 \leq r_1 \leq 1$; FT —in the current sequence position j , the tasks that ant can be selected with the earliest start time; α , β —parameters determine the relative importance of pheromone intensity and heuristic information; η_i —the heuristic information of task i , using pw_i that is the position weight of task i as heuristic information:

$$\eta_i = pw_i = t_i + \sum_{j \in F_i} t_j, \quad i = 1, \dots, N \quad (11)$$

F_i —set of successor tasks of task i . Position weight heuristic rules account the sum of successor task operating time and the largest number of successor tasks

simultaneity, tasks with largest successor task numbers and longer sum operating time of successor tasks has the greater probability of being select. In formulation (10), the random variable r determines using the rules below to select a task form FT : If the random number r satisfies $0 \leq r \leq r_1$, then randomly select a task from the FT with the probability p_{ij} ; if the random number r satisfies $r_1 < r \leq 1$, then randomly choose a task from the candidate FT .

3.2 Heuristic Task Distribution Rules

Rule 1: If the task’s operating position constraint is L (or R), it is placed on the left (or right) station; Rule 2: If this task’s operating position constraint is E , then put it into the side which it can began earlier; if it placed on both sides with the same start time, then selected one side randomly.

3.3 Construct the Feasible Solution in Ant Colony Algorithm

As shown in Fig. 2, the open position is the determined value n , and the optimization goal is to minimize the cycle time. In order to ensure that all tasks can be completely distributed, in the first $n - 1$ positions, the constraints of the optimal cycle time are used; in the final position of n , eliminating the cycle time constraint to assign all tasks into this location to meet the total number of the open position is exactly n , and feasible solutions are generated. Find out the latest completion time from all positions in this feasible solution as the cycle time searched by this ant.

3.4 Pheromone Updating Rule

Local pheromone update: The role is to reduce the attractive effect from earlier ants to the later one, thereby this method strengthens the random search capabilities of ants. When the ant assigned task i to the arranged location j , the pheromone in edge ij updated as formulation (12):

$$\tau_{ij} \leftarrow (1 - \rho_1)\tau_{ij} + \rho_1\tau_0 \tag{12}$$

ρ_1 local pheromone evaporation coefficient, $0 < \rho_1 < 1$;

τ_0 initialized pheromone value, $\tau_0 = 1/(N \cdot K^*)$, $K^* = \left[\left(\sum_{i \in I} t_i \right) / C_{\min} \right]$.

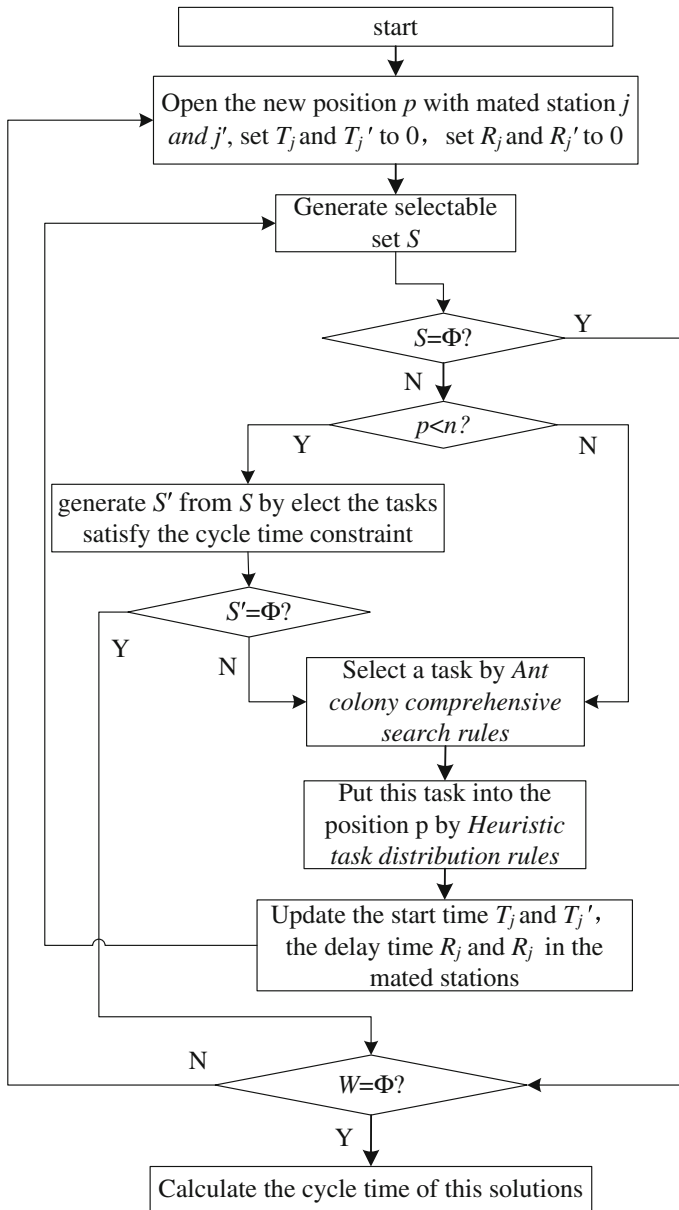


Fig. 2 Assignment of tasks to construct an ant solution

Global pheromone update rule: As the establishment of an optimal solution set for each generation of ants' search results in this paper, every optimal solution in

the collection releases pheromone to the path. The global pheromone update formula is as follows:

$$\tau_{ij} \leftarrow (1 - \rho_2)\tau_{ij} + \rho_2\Delta\tau_{ij} \tag{13}$$

where

$$\Delta\tau_{ij}^{gb} = \begin{cases} 1/N_S, & \text{if } (i, j) \in \text{global-best-tour} \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

ρ_2 is global pheromone evaporation coefficient, $0 \leq \rho_2 \leq 1$

The whole search plan of ant colony algorithm: The algorithm starts search the optimal solution at the theoretical minimum cycle time, C_{min} . The first generation of ants search from the smallest cycle time C_t ($C_t = C_{min}$) at the start search. If it do not find the solution of C_t , the standard value C_t increased by 1 in the next generation of ant colony search, $C_t = C_t + 1$; if a generation of ant colony obtained the current search criteria C_t , then the next generation of search will narrow the search criteria, that is, $C_t = C_t - 1$; the entire ant colony uses this plan to search for the theoretical optimal solution until find out the optimal solution or the maximum search iteration is reached.

4 Computational Study

Precedence diagram of a loader assembly line [11] is given in Fig. 3. The proposed ant colony algorithm was coded in MATLAB 7.8 and run on a Notebook PC equipped with Intel Core i5-2410 M 2.30 GHz and 4 GB memory. In the proposed ACO, several parameters, such as number of ants n_{ant} , maximize iteration number NC_{max} , α , β , ρ and e , affect the performance of the algorithm. Excessive experimental tests were conducted and found out the following value of parameters are more appropriate: $n_{ant} = N$, $NC_{max} = 400$, $\alpha = 1$, $\beta = 2$, $\rho_1 = 0.1$, $\rho_2 = 0.1$, $r_1 = 0.8$. By changing the number n (number of opened positions in the assembly line), the computational study shows the four kinds of optimization solutions as shown in Table 1. Figure 4 is the Gantt chart output by MATLAB as the second optimization solution in Table 1.

In Table 1, four kinds of optimization results are all with a good balance rate. Assignment schedules of scheme 2 for the test problem generated by ACO refer to Table 2. To increase the practicality of the algorithm and the intuitive of the results, the Gantt chart display module was added to this program, refer to Fig. 3. The progress bar's length is in accordance with the tasks' operation time, and the number on top of the progress bar is the start time and end time of tasks; the horizontal axis is for the timeline for the task start and end time, the vertical axis is for the station, which on the left station of every position is odd, the right station in

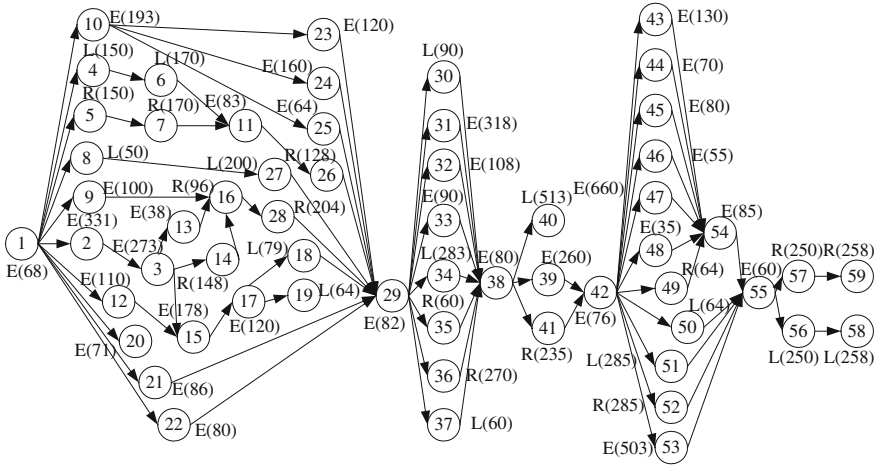


Fig. 3 Precedence diagram of a loader assembly line

Table 1 Four types of assignments for the test problem generated by ACO

Scheme	C(s)	LE (%)	SI (s)
Scheme 1(5)	1,021	94.0	79.73
Scheme 2(6)	860	93.0	70.22
Scheme 3(7)	771	89.0	152.47
Scheme 4(8)	660	91.0	81.01

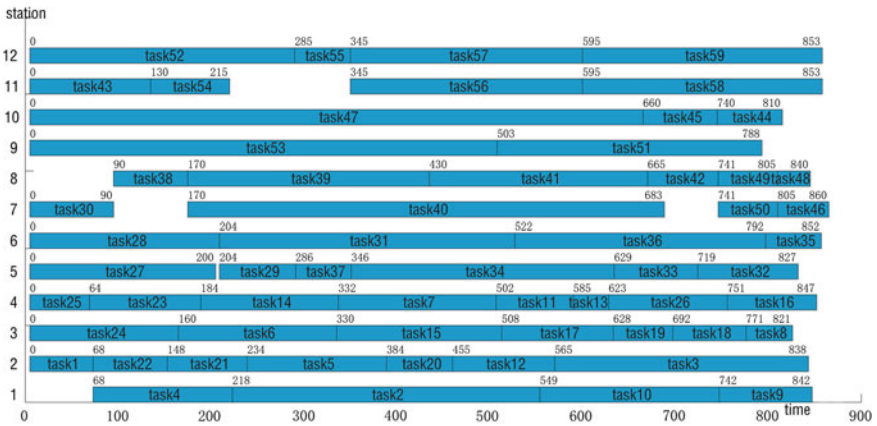


Fig. 4 Gantt chart of assignment of tasks for the example problem

Table 2 Assignment schedules of scheme 2

Work station	Assigned tasks	Station (waiting) time (sec)
1 (1L)	4, 2, 10, 9	774 (68)
2 (1R)	1, 22, 21, 5, 20, 12, 3	838
3 (2L)	24, 6, 15, 17, 19, 18, 8	821
4 (2R)	25, 23, 14, 7, 11, 13, 26, 16	847
5 (3L)	27, 29, 37, 34, 33, 32	823 (4)
6 (3R)	28, 31, 36, 35	852
7 (4L)	30, 40, 50, 46	722 (138)
8 (4R)	38, 39, 41, 42, 49, 48	750 (90)
9 (5L)	53, 51	788
10 (5R)	47, 45, 44	810
11 (6L)	43, 54, 56, 58	723 (130)
12 (6R)	52, 55, 57, 59	853

the same position is even. The time delayed by the pre-order tasks in the mated-station is clear in Fig. 1, so the Gantt chart display module has obvious practical significance.

5 Conclusions

In this paper, an ant colony algorithm was introduced to solve two-sided assembly line balance problem type-II. An ant colony search rules, heuristic task distribution rules were used for generate a feasible solution. The whole search plan of ant colony algorithm controls the optimal solution search process. Computational experiment demonstrated the validity of the proposed algorithm. In the future studies, we hope to apply the method to extensions of the TALBP-II, such as multiple-object TALBP-II.

Acknowledgments This research was partially supported by the National Natural Science Foundation of China (No. 51205328), the Youth Foundation for Humanities and Social Sciences of Ministry of Education of China (No. 12YJJCZH296), the PhD Programs Foundation of Ministry of Education of China (No. 200806131014), the Foundation of Sichuan Province Cyclic Economy Research Center (No. XHJJ-1205), and the Fundamental Research Funds for the Central Universities (No. SWJTU09CX022; 2010ZT03).

References

1. Bartholdi JJ (1993) Balancing two-sided assembly lines: a case study. *Int J Prod Res* 31:2447–2461
2. Kim YK, Kim YH, Kim YJ (2000) Two-sided assembly line balancing: a genetic algorithm approach. *Prod Plan Control* 11:44–53

3. Lee TO, Kim Y, Kim YK (2001) Two-sided assembly line balancing to maximize work relatedness and slackness. *Comput Ind Eng* 40:273–292
4. Simaria AS, Vilarinho PM (2009) 2-ANTBAL: an ant colony optimisation algorithm for balancing two-sided assembly lines. *Comput Ind Eng* 56:489–506
5. Baykasoglu A, Dereli T (2008) Two-sided assembly line balancing using an ant-colony-based heuristic. *Int J Adv Manuf Technol* 36:582–588
6. Wu EF, Jin Y, Bao JS, Hu XF (2008) A branch-and-bound algorithm for two-sided assembly line balancing. *Int J Adv Manuf Technol* 39:1009–1015
7. Scholl A, Becker C (2006) State-of-the-art exact and heuristic solution procedures for simple assembly line balancing. *Eur J Oper Res* 168:666–693
8. Ozcan K (2010) A Petri net-based heuristic for simple assembly line balancing problem of type 2. *Int J Adv Manuf Technol* 46:329–338
9. Kim YK, Song WS, Kim JH (2009) A mathematical model and a genetic algorithm for two-sided assembly line balancing. *Comput Oper Res* 36:853–865
10. Zhang Z-Q, Cheng W-M, Zhong B, Wang J-N (2007) Improved ant colony optimization for assembly line balancing problem. *Comput Integr Manuf Syst CIMS* 13:1632–1638
11. Wu E (2009) Research on balancing two-sided assembly line. School of mechanical engineering, vol 120. Doctor of Science. Shanghai Jiao Tong University, Shanghai (2009)

Detection of Earnings Manipulation with Multiple Fuzzy Rules

Shuangjie Li and Hongxu Liang

Abstract Using a multi-objective linear programming, we develop an approach to set the decision power for the multiple fuzzy rules, properly in a highly fuzziness event and earnings manipulation, whose degree of membership is difficult to observe. We use the proposed method to detect the uncovered earnings manipulators during the period 2001–2010 from the companies listed at Shanghai Stock Exchange and Shenzhen Stock Exchange. The recognition rate for in-sample test is 78.4 %, and the corresponding rate for out-of-sample test is 76.9 %.

Keywords Earnings manipulation · Fuzzy rules · Multi-objective linear programming

1 Introduction

The study of detecting earnings manipulation starts from the research of earnings management in 1980s. Healy (1985) first raises a quantized model to evaluate the level of the existence of earnings management, which is also the beginning of quantitative analysis on the research of earnings manipulation. The managers' initial purpose of earnings management should only to keep their interests on achieving stable, comparable, and predictable financial results legally, so that the financial reports could give a better reflection of the company's economic reality to the users. But for some reasons, the generally accepted accounting principles

S. Li (✉) · H. Liang

School of Economics and Management, Beijing University of Technology,
Pingleyuan 100, 100124 Beijing, China
e-mail: lishuangjie@bjut.edu.cn

H. Liang

e-mail: lianghxic@sina.com

(GAAP), which is the bottom line of earnings management, would be explicated in a radical way and the financial statements would fail to reflect the economic reality and mislead the users. While this happens, people also called this unethical or illegal activity as earnings manipulation or “cooking the books” instead of earnings management. In this paper, we focus on the companies that reported to have more income than their real income, since this kind of earnings manipulation could bring much more loss to the investors compared with the other kinds of earnings manipulation, which tends to report less income.

Although the regulatory measures for the earnings manipulation have been continuously strengthening, some public companies, including the industrial giants such as Enron, WorldCom, and OLYMPUS, still could easily evade supervision in the first few years. Thus, earnings manipulation may not vanish if we only rely on the increasing penalties and auditing, because they could hardly give an effective recognizing or early warning until the earnings manipulation is unavoidable. For this reason, some researchers put their interests in detecting the earnings manipulation with statistical and econometric models.

However, since the earnings manipulation is a continuous behavior and the recognition rules are always linguistic information, in this paper, we propose a new model to deal with the fuzzy set theory (FST) and the linear programming (LP). In Sect. 2, we recall some researches on the earnings manipulation briefly. The new discriminate model would be introduced in Sect. 3. In Sect. 4, we would show an empirical study of the detection of Chinese public companies’ earnings manipulation based on the model introduced in Sect. 3. At last, we provide a conclusion.

2 A Brief Review of the Research on the Earnings Manipulation

So far, some frequently used methods to detect the earnings manipulation could briefly grouped as follows: (a) Jones and modified Jones model: Chou et al. [1] study the relationship between the earnings management and the stock performance after the reverse leveraged buyouts with Jones Model; Abarbanell and Lehavy [2] use Jones model to show that the stock recommendations could be used to predict the earnings management; Phillips et al. [3] find that the deferred tax expense could be used to detect the earnings management after comparing the regression results from modified Jones model and forward-looking model, and the usefulness of deferred tax expense is supported while using to detect earnings management; Zhang [4] found that there is earnings management for some of the companies in the year before private placement in China with modified Jones model. (b) Probit or logit model: Beneish [5] proposed a probit model to detect the earnings manipulation which could detect half of the manipulators before they were uncovered; Jiang et al. [6] proposed a logit model to detect the earnings manipulation of Chinese public companies, of which 70–80 % of the manipulators

could be detected based on the paired samples; Barton and Simko [7] found that the level of net assets could be seen as a constraint of earnings management, and before estimating the logit model, they measure the overstatement in net asset values by Jones model. (c) Statistical analysis approaches: Li et al. [8] detect the manipulators in China through the Bayes classification analysis, and the classification model's detection power reached about 70 % for the paired sample of manipulators and normal companies in the empirical study; Yao et al. [9] detect the earnings manipulation with the principal component analysis (PCA).

All the methods we aforementioned considered the earnings manipulators and the normal companies as a binary value, like boolean or logical variables, to simply symbol the level or possibility of earnings manipulation that they could not be observed directly. However, this simplified explanation for the difference between the earnings manipulation and the common operation behaviors such as earnings management could have a big problem. For there has no distinct boundary between the earnings manipulation and the earnings management, it is a continuous process rather than a discrete process. From both surveys from *CFO magazine* and *Business Week*, it shows that the earnings management is broadly occurred in the normal companies as a common operating decision. As a result, it is not very proper to divide the manipulators and the normal companies into two separate groups; those companies in the same group may be considered to have some same levels of risk; the members of group A may be seen as safe or AAA level, while the members of group B may be seen as risk or CCC level. This information could mislead the users seriously, especially for those companies which should belong to the invisible group "middle" or "neutral." At the same time, in reality, the advices that investors could get to detect the existence of earnings manipulation from the analysts and researchers are always some linguistic discriminant rules, such as "If indicator A is too large, Then this company could have a higher level of risk," and it is hard for the users to determine the decision power of those rules. Therefore, in this paper, we will use the fuzzy set theory to deal with this continuous process and those linguistic rules.

3 The Detection of Earnings Manipulation

For the normal companies, there always has some degree of earnings management, since the GAAP give some flexible spaces to accountant for some accounting items; the managers would use some of the discretion subjectively or objectively, while they exceed the invisible boundary too much, maybe for avoiding a loss, avoiding an earnings that decline or fail to meet the analysts' forecasts, the minor fault of ethics would turn into the fraud, and the earnings management then becomes the earnings manipulation. And whether the behavior is serious enough to be affirmed as the earnings manipulator partly depends on the judgment of regulators. In order to handle the events that the human logical thinking takes part and without a clear demarcation, Zadeh (1965) first introduces the concept of "fuzzy

set”, so that we could use a fuzzy number to represent the human thinking approximately. The main difference between FST and set theory is that the former admits the nature of things that could have a “transition state,” while the other believes that everything could be grouped by some different crisp subsets which have an empty intersection. The relationship between these theories is not exclusive but complementary in the practical application.

Before we construct the model, we need to add a weak assumption: If there exists a set of indicators which could detect the earnings manipulation well, for those manipulators the degree of membership for earnings manipulation should be higher than that of earnings management or normal accounting operation. This assumption is just a natural extension of maximum membership principle, and it allows us to simulate the human thinking process, and then, we could get the estimation that converges to the analysts’ feeling as well as possible, even though we could not observe the exact real degree of membership.

First, we need to select some indicators or ratios that are significant while used to detect the earnings manipulation. For some financial indicators, it could not be seen as asymptotically normal and the kind of distribution is hard to be determined. Moreover, the same indicator from different industries or fiscal years is not comparable. To overcome this dilemma, we use the quantile to make the indicator comparable. This means we could sort the data of manipulators with the normal companies that come from the same industry and same year, the percentile we get is just the relative level of the company for this indicator, it could tell us how large or small this indicator is regardless of the industry and fiscal year. Obviously for all the companies, the set of percentile should be uniformly distributed, and then, we could estimate the means of the percentile of manipulators for each indicator to test the significance while used to detect the earnings manipulation. Since the ratio of manipulators is relatively low for the entire public companies, the set of percentiles for the normal companies should be asymptotically distributed as $U(0, 1)$. Under the null hypothesis, which means that there has no significant difference between manipulators and normal companies for the indicator, the percentiles of manipulators should also be uniformly distributed. Furthermore, it is independently and identically distributed (i.i.d) with mean 0.5 and covariance 1/12. With Lindeberg–Lévy central limit theorem, under null hypothesis, the mean of manipulators’ percentile should follow

$$\sqrt{n} \left(\left(\frac{1}{n} \sum_{i=1}^n X_i \right) - \mu \right) \xrightarrow{d} N(0, \sigma^2) \quad (1)$$

Otherwise, we could believe that this indicator should have some power to detect the earnings manipulation. Since there has no explicit logical relationship between those indicators, we could hardly simply use the “or,” “and,” and “not” to handle these complex phenomena. In this paper, we do as C-Tsao [10] which set a decision power or weight to each fuzzy number, but in C-Tsao’s paper, there did not give a common way to estimate the weights. In our paper, we use a multi-

objective linear programming model to estimate the decision power of each indicator.

While we use the FST in practice, one of the most important questions is how to define the membership function properly. There are many approaches to be chosen, but there is no approach absolutely the “best” one once and for all. Just like Pietraszek and Dwornicka [11] said, some “convenient” form of membership function, such as the triangle and the trapezoid, may probably different from the real circumstances. To avoid the subjective effect bias in the definition of membership function, and to maintain the ordering structure of the original data, we define the membership function based on the direct transform of cumulative distribution function (CDF). Although there is big difference between membership function and CDF, in recent years, more and more researchers found that there is some dual relationship between these two concepts in some cases. Guo and Love [12] point that while the membership function is known as monotonic, some membership functions may be the direct transform of CDFs, and the percentile may contain some key information about membership function. Then, Guo [13] defines the membership function based on the percentile. Teoh et al. [14] define the membership functions of some stock market indices through the CDF. Fal-safain et al. [15] define the membership function based on the CDF, while they use fuzzy method to estimate the parameters in statistical models. Chen et al. [16] define the membership function based on the transform of the CDF of normal distribution.

We define the membership as follows: for the indicator when its level becomes higher that could lead to a larger possibility in earnings manipulation, the degree of membership for the normal earnings management or normal financial operation could be defined as

$$\begin{aligned} \mu_N(x_{0k}) &= 1 - F_N(x_{0k}) = 1 - P(v \leq x_0 | v \in N_{\bullet k}) \\ &= 1 - \int_{v=0}^{x_{0k}} f(v)dv = 1 - x_{0k} \end{aligned} \tag{2}$$

where x_{0k} is the quantile of the company which would be detected for the k th indicator and $N_{\bullet k}$ is the set of the quantiles of normal companies for the k th indicator. The last part of equation comes from the fact that the percentiles of the normal companies are distributed as uniform distribution asymptotically. Similarly, we can define the degree of membership for the earnings manipulation as

$$\begin{aligned} \mu_M(x_{0k}) &= F_M(x_{0k}) = P(v \leq x_{0k} | v \in M_{\bullet k}) = \sum_{i=1}^n B_i/n \\ \text{where } B_i &= 1, \quad v_i \leq x_{0k}; \\ &= 0, \quad \text{otherwise} \end{aligned} \tag{3}$$

$M_{\bullet k}$ is the set of quantiles of all the manipulators compared with the normal companies from the same industry and year correspondingly for the k th indicator.

n is the total number of the manipulators uncovered. The right-hand side of the equation comes from the fact that the manipulators' quantile is a discrete distribution.

We could define the indicator which tends to be more likely to have earnings manipulation, while the quantile decreases. The degree of membership for the legal earnings management could be defined as

$$\mu_N(x_{0k}) = F_N(x_{0k}) = P(v \leq x_{0k} | v \in N_{\bullet k}) = \int_{v=0}^{x_{0k}} f(v)dv = x_{0k} \quad (4)$$

The degree of earnings manipulation could be defined as

$$\begin{aligned} \mu_M(x_{0k}) &= 1 - F_M(x_{0k}) = 1 - P(v \leq x_{0k} | v \in M_{\bullet k}) \\ &= 1 - \sum_{i=1}^n B_i/n \end{aligned} \quad (5)$$

Obviously, for most of the normal companies, the degree of membership for normal company should be larger than that for manipulators.

As we mentioned before, there has no effective way to estimate the real degree of membership of the uncovered manipulator, neither how likely may it belong to the manipulator nor how likely may it belong to the normal company could be predetermined. However, we could still estimate the difference between the degree of membership for the normal companies and the degree of membership for the manipulators, and the decision power of the indicators or fuzzy rules simultaneously with the assumption we have made in Sect. 3 through a multi-objective integer LP (ILP) model.

The first objective is to find some weight vectors which could detect the earnings manipulation from the manipulators' sample as well as possible, that is, to minimize the number of detection failure. Then, the step 1 of the model could be written as

$$\begin{aligned} \text{Min} \quad & \sum_{i=1}^n d_i \\ \text{s. t. :} \quad & \sum_{r=1}^s w_r(m_{ri} - n_{ri}) + d_i \geq 0 \quad i \in M \\ & d_i = \{0, 1\} \\ & \sum_{r=1}^s w_r = 1 \end{aligned} \quad (6)$$

where M denotes all the manipulators we selected, s is the number of the fuzzy rules (indicators), and m and n denote the degree of membership for the manipulators and normal companies correspondingly. Obviously, our principle of detecting this model is to compare the difference between the degree of membership for manipulators and normal companies.

The second step is to find a weight vector from step 1, which could minimize the total detecting error. Since this weight vector is selected from step 1, the detection ratio would be equal to the ratio in step 1.

$$\begin{aligned}
 & \text{Min } \sum_{i=1}^n t_i \\
 & s.t. : \sum_{r=1}^s w_r(m_{ri} - n_{ri}) + t_i \geq 0 \quad i \in M \\
 & \quad t_i \leq d_i \\
 & \quad d_i = \{0, 1\} \\
 & \quad \sum_{i=1}^n d_i = [\text{obj}_{\text{step 1}}] \\
 & \quad \sum_{r=1}^s w_r = 1 \\
 & \quad t_i \geq 0
 \end{aligned} \tag{7}$$

The sample including all the manipulators uncovered by the regulator whose earnings manipulation had continuously occurred more than 3 years. To make sure that every manipulator’s information could have an equal effect to the objective, the data we selected only contain the first 3 years after the earnings manipulation first occurred. For the manipulators that made, earnings manipulation less than 3 years would be selected for out-of-sample test. After we run this model, we could get the decision power of each fuzzy rule, and then, we could estimate the possibility of occurring earnings manipulation for any company with the maximum membership principle.

4 Empirical Results

While we select the indicators based on the fuzzy rules mentioned by other researchers and analysts, we need to consider the effect of the Chinese new accounting standards carried out in 2007. It makes the definitions of some indicators changed a lot, which means it may become incomparable for the data before 2007 and after 2007. As a result, we get the adjusted data from the CCER database, and it makes most indicators comparable. We will not select those indicators changed too much, so that we could make sure that the change in indicators will not change the fuzzy rules’ linguistic meaning significantly.

The financial data of all the public companies used in this study are obtained from the CCER database. We collect the uncovered manipulators which had the occurrence of earnings manipulation during 2001 to 2010. As we stated in the previous section, the data of manipulators we choose at most contain the first

Table 1 The indicators and the means of quantiles

No	Variable in model	Indicator	Mean of quantiles
1	ITR	Inventory turnover ratio	0.4387
2	ADI	Accounts receivable/net sales	0.6486
3	OAI	Other receivables/net sales	0.6716
4	PDA	Prepayment/total assets	0.6124
5	PEA	Net property and equipment/total assets	0.396
6	PDC	Period cost/cost of sales	0.644
7	RDI	Operating income/net sales	0.3709
8	CAI	Total current assets/net sales	0.6624
9	CNI	Net cash provided by operations/net sales	0.4214
10	OCA	Cash received relating to other operating activities/cash inflows from operating activities	0.6234

3 years after the earnings manipulation first occurred; after we omit the missing data, there left 128 sets of financial data from the fiscal year 2001–2010.

Before we run the model, we need to determine which fuzzy rules that the analysts and researchers have mentioned should be chosen as a detection indicator for Chinese stock market. From (1), we could do it by testing the mean of quantiles with a t test under the null hypothesis, and the corresponding confidence interval is $[0.45, 0.55]$ with the significant level of 5%. After we get the set of indicators rejected by the null hypothesis, we must estimate the Spearman's rho between these indicators. If there are some indicators which have relatively large Spearman's correlation coefficients that could cause serious bias while we estimate the decision power for these fuzzy rules, it could mean that there may have a collinear problem while we run the model. When this happened, we have to decide which indicator should left in the set and which should be omitted. The indicators we choose for the model are shown in Table 1.

After we run the model with the indicators presented in Table 1, the decision powers of these indicators could be estimated. The weights of the indicators are shown in Table 2.

The objective value of step 1 which shows the number of company that failed to be detected is 22. There are totally 102 (3 fiscal years*34 manipulators) groups of the financial data used in the model. The ratio of detection is 78.4%. The objective value of step 2 is 5.582647, which is the minimized sum of deviation from those individuals failed to be detected.

We also make an out-of-sample test with the weight vector which is shown in Table 2 for the remaining 26 records from 15 manipulators which manipulated the earnings less than 3 years. Twenty of them have been found that the degree of membership for the earnings manipulators is larger than that for the normal companies, which means we considered these 20 records would have a high level of risk for investors. The ratio of detecting the earnings manipulation correctly arrives at 76.92%. This ratio of out-of-sample test is almost the same with the result from the in-sample data. It shows that the detection power of model is asymptotically consistent.

Table 2 The decision power of indicators

Indicator	Decision power	Indicator	Decision power
ITR	3.845 E-05	PDC	0.005013
ADI	0.23036	RDI	0.058518
OAI	0.052182	CAI	0
PDA	0.21303	CNI	0.093395
PEA	0.214855	OCA	0.132606

5 Conclusion

In this paper, we discussed the detection of earnings manipulation for the companies listed in Chinese stock market. Firstly, we select the indicators based on the average level of quantiles, which is the indicator’s relative level in ascending order in the same industry and same fiscal year for each manipulator. Some of the indicators which have higher Spearman’s correlation coefficients with other indicators should be removed. Then, we transform these comparable quantiles properly to define the membership function for both the activities of earnings manipulation and the normal financial operation. Since the earnings manipulation is a highly complex economic behavior, there is no direct approach to predetermine company’s exact degree of membership with the information set composed by the data given by the financial statements, and we could hardly found distinct logical relationships between the multiple fuzzy rules that the analysts or researchers have given. In this paper, we propose a multi-objective LP model to handle these two problems simultaneously based on the principle of maximum membership. After we run the model, we could estimate the decision powers for each indicator and we could also get each company’s difference between the degree of membership for manipulators and normal companies, so that we could determine whether this company has a relatively high risk for the investors. The result of in-sample test and the out-of-sample test for the detection power is very similar, and it shows that this model could detect more than 3/4 of the manipulators. And this model could give users more detailed information about the existence of earnings manipulation. They could calculate the degree of membership with weights the model has given. They could see how “far” the company would become to the manipulator even if they could see it as a normal company now, and vice versa, while the traditional detecting models only could tell the users an 0 or 1 result for the companies on the brink of critical point.

As we stated before, we could not use some of fuzzy rules for the implementation of new accounting standards, especially for the indicators which include the ratios between two different fiscal years. This may reduce the model’s detecting power to some degree. Moreover, further research is concerned to add some non-monotonic fuzzy rules as the supplement for the model.

References

1. Chou DW, Gombola M, Liu FY (2006) Earnings management and stock performance of reverse leveraged buyouts. *J Fin Quant Anal* 41(2):407–438
2. Abarbanell J, Lehavy R (2003) Can stock recommendations predict earnings management and analysts' earnings forecast errors? *J Account Res* 41(1):1–31
3. Phillips JJ, Pincus M, Rego SO (2003) Earnings management: new evidence based on deferred tax expense. *Account Rev* 78(2):491–521
4. Zhang WD (2010) Private placement of new shares and earnings management—Empirical evidence from China's Stock market. *Manag World* 1:54–63
5. Beneish MD (1999) The detection of earnings manipulation. *Fin Anal J* 55(5):24–36
6. Jiang JL, Li YX, Gao R (1999) Detecting earnings manipulation of listed companies based on logistic model. *Econ Manage J* 30:24–36
7. Barton J, Simko PJ (2002) The balance sheet as an earnings management constraint. *Account Rev* 77:1–27
8. Li YX, Gao R, Bao SZ, Yao H (2007) Bayes classifying analysis of earnings manipulation in listed companies in China. *Forecasting* 3:56–60
9. Yao H, Li YX, Gao R (2007) A recognition model of aggressive earnings management in Chinese listed companies based on the principal components method. *J Manage Sci* 20:83–91
10. Tsao CT (2006) A fuzzy MCDM approach for stock selection. *J Oper Res Soc* 57(11):1341–1352
11. Pietraszek J, Dwornicka R (2010) Estimation of accuracy for empirical assessment of membership function's value. *Czasopismo Techniczne Mechanika* 107(8):219–228
12. Guo RK, Love CE (2003) Reliability Modeling with Fuzzy Covariates. *Int J Reliab Qual Safety Eng* 10(2):131–157
13. Guo RK (2005) A grey semi-statistical fuzzy modelling of an imperfectly repaired system. In: *Proceedings of the 4th international conference on quality and reliability (ICQR 2005)*, pp 391–399 (2005)
14. Teoh HJ, Cheng CH, Chu HH, Chen JS (2008) Fuzzy time series model based on probabilistic approach and rough set rule induction for empirical research in stock markets. *Data Knowl Eng* 67:103–117
15. Falsafain A, Taheri SM, Mashinchi M (2008) Fuzzy estimation of parameters in statistical models. *Int J Comput Math Sci* 2(2):79–85
16. Chen JS, Chou HL, Cheng CH, Wang JY (2011) CPDA based fuzzy association rules for learning achievement mining. In: *2009 international conference on machine learning and computing, IPCSIT*, vol. 3. IACSIT Press, pp 25–29

A New Unbalanced Linguistic Scale for the Classification of Olive Oil Based on the Fuzzy Linguistic Approach

M. Espinilla, F. J. Estrella and L. Martínez

Abstract A key factor that determines the price of olive oil is its sensory profile. The International Olive Council (IOC) establishes four quality categories and a method to classify a sample of olive oil into one category, depending on its sensory characteristics. To do so, a taster panel is rigorously trained to provide the intensity perceived on a 10-cm scale for each organoleptic characteristic. These intensities are aggregated and analyzed statistically to obtain the classification among one of four quality categories established. The modeling and management of perceptions in sensory evaluation processes is an important problem because the information acquired by human senses always involves imprecision and uncertainty that has a non-probabilistic nature. The application of the fuzzy linguistic approach to sensory evaluation processes can model and manage the uncertainty and vagueness of this kind of processes. The main challenge in this approach is to establish a linguistic scale to measure tasters' perceptions, since the success or failure of the sensory evaluation process will depend on the definition of a proper scale. In this contribution is analyzed and proposed an unbalanced linguistic scale to carry out the classification of olive oil samples, such a scale is validated, conducting a sensory evaluation case study for olive oil.

Keywords Sensory evaluation · Fuzzy linguistic approach · Unbalanced linguistic scale · Olive oil · Linguistic 2-tuple.

M. Espinilla (✉) · F. J. Estrella · L. Martínez
Department of Computer Sciences, University of Jaen, 23071 Jaén, Spain
e-mail: mestevez@ujaen.es

F. J. Estrella
e-mail: estrella@ujaen.es

L. Martínez
e-mail: martin@ujaen.es

1 Introduction

Sensory evaluation is an evaluation discipline in which the information provided by a panel of individuals is perceived by human senses of *sight, smell, taste, touch, and hearing*. This evaluation is generally applied to the quality assurance for products, to solve conflicts between customers and producers, to develop new products, and to exploit new markets adapted to the consumer's preference [1, 2].

The sensory information like color, flavor, taste, and mouthfeel are generally obtained through subjective information (perceptions). This fact implies the following main difficulties in sensory evaluation processes:

- *D1*: The information presented in a sensory evaluation process always implies uncertainty and imprecision which are generally analyzed statistically, assuming that any uncertainty can be represented by a probabilistic distribution. However, this information has a non-probabilistic nature [3].
- *D2*: One key issue for the success of a sensory evaluation process depends on the correct definition of the scale used to measure the sensory information. This definition is not trivial because it requires to fix the structure of the scale, the number of the terms, its distribution, etc.

Researches have shown that the fuzzy linguistic approach [4] and the fuzzy set theory [5] are considered useful tools to model and manage the uncertainty in sensory evaluation processes of many products [3, 6, 7] like mango drink [8], tea [9], coffee [10], sausages [11], or Indian yogurt [12]. So, the proposed linguistic sensory evaluation model [13] for olive oil overcomes the first difficulty, by using the fuzzy linguistic approach [4] to model and manage such an uncertainty.

In this contribution, we are focused on the quality of olive oil because this is a key factor in its marketing: *an excellent quality implies a higher price in the market* [14, 15]. The quality of a sample of olive oil is established by its sensory profile in which each sensory attribute (positive or negative) is measured by a trained tasters' panel. The International Olive Council (IOC) establishes four quality categories for the olive oil: *virgin extra, virgin, ordinary, and lampante* and fixes the procedure to assess the organoleptic characteristics and classify the olive oil on the basis of these characteristics. So, each taster provides the intensity perceived of each attribute on a 10-cm scale in a profile sheet. The olive oil is categorized, taking into account the median value of the negative attributes and the median for the fruity attribute (positive attribute), according to reference ranges: 0, 3.5, and 6. A detailed description about the procedure and attributes can be found in IOC/T.20/Doc. No 15/Rev. November 4, 2011.¹

In a previous work, [13] was proposed a linguistic sensory evaluation model to establish the category of a sample of olive oil, dealing with an unbalanced linguistic scale [16, 17]. The scale proposed was a five-term scale whose distributions were defined according to the reference ranges to classify the olive oil. However,

¹ www.internationaloliveoil.org/documents/viewfile/3685-orga6

recently, we have detected that such an unbalanced linguistic scale fails in the difficulty *D2* because samples of olive oil are classified incorrectly when olive oils are doubtful between two categories.

The aim of this contribution is to analyze with two taster panels of olive oil an adequate unbalanced linguistic scale and then to validate it, carrying out a sensory evaluation case study for a set of samples of olive oil, belonging to different categories.

The rest of the contribution is set out as follows. [Section 2](#) reviews some linguistic concepts necessary to understand our proposal. [Section 3](#) introduces in short the unbalanced linguistic sensory evaluation model utilized in our sensory evaluation case study. [Section 4](#) presents in detail the proposal of the unbalanced linguistic scale and its validation, carrying out a sensory evaluation case study. Finally, in [Sect. 5](#), conclusions are drawn.

2 Linguistic Background

Due to the use of linguistic information and processes of computing with words [18] in the olive oil evaluation, here we review some concepts used.

2.1 Fuzzy Linguistic Approach

Sensory information is the information perceived by the human senses of *sight*, *smell*, *taste*, *touch*, and *hearing*. This information implies uncertainty, vagueness, and imprecision, and the use of the fuzzy linguistic approach [4] has provided successful results modeling this kind of information. The fuzzy linguistic approach represents this information as linguistic values by means of linguistic variables [4]. Usually, in these cases, it is required that in the linguistic term set, there exist:

1. A negation operator: $\text{Neg}(s_i) = s_j$ such that $j = g - i$ ($g + 1$ is the cardinality).
2. An order: $s_i \leq s_j \iff i \leq j$. Therefore, there exists a *min* operator and a *max* operator.

The semantics of the terms are given by fuzzy numbers defined in the $[0,1]$ interval, which are usually described by membership functions.

2.2 2-Tuple Linguistic Representation Model

The use of linguistic information implies to operate with such a type of information, that is, processes of computing with words (CWs). In [19] was presented a linguistic representation model based on linguistic 2-tuples that carries out

processes of CW in a precise way when the linguistic term sets are symmetrical and uniformly distributed.

The linguistic 2-tuple representation model is based on the concept of *symbolic translation* [19] and represents the linguistic information through a 2-tuple (s, α) , where $s \in S = \{s_0, \dots, s_g\}$ is a linguistic term and α is a numerical value representation of the symbolic translation [19]. Thereby, being $\beta \in [0, g]$ the value generated by a symbolic aggregation operation, we can assign a 2-tuple (s, α) that expresses the equivalent information of that given by β .

Definition 1 [19]. Let $S = \{s_0, \dots, s_g\}$ be a set of linguistic terms. The 2-tuple set associated with S is defined as $\langle S \rangle = S \times [-0.5, 0.5]$. We define the function $\Delta_S : [0, g] \rightarrow \langle S \rangle$ given by

$$\Delta_S(\beta) = (s_i, \alpha), \text{ with } \begin{cases} i = \text{round}(\beta), \\ \alpha = \beta - i, \end{cases} \tag{1}$$

where *round* assigns to β the integer number $i \in \{0, 1, \dots, g\}$ closest to β .

We note that Δ_S is bijective [19], and $\Delta_S^{-1} : \langle S \rangle \rightarrow [0, g]$ is defined by $\Delta_S^{-1}(s_i, \alpha) = i + \alpha$. In this way, the 2-tuples of $\langle S \rangle$ will be identified with the numerical values in the interval $[0, g]$.

The linguistic 2-tuple representation model has a linguistic computing model associated that accomplishes CW processes in a precise way. Different aggregation operators have been proposed for linguistic 2-tuple [19–22]. In our proposal, we will use the median aggregation operator for linguistic 2-tuple since the IOC computes collective sensory intensities based on the calculation of their medians.

Definition 2 [13]. Let $((s_1, \alpha), \dots, (s_n, \alpha)) \in \langle S \rangle^n$ be a vector of linguistic 2-tuples. The 2-tuple median operator is the function $\text{Med} : \langle S \rangle^n \rightarrow \langle S \rangle$ defined by if *n* is odd

$$\text{Med}((s_1, \alpha), \dots, (s_n, \alpha)) = (s_i, \alpha)$$

if *n* is even

$$\text{Med}((s_1, \alpha), \dots, (s_n, \alpha)) = \Delta_S\left(\frac{\Delta_S^{-1}(s_i, \alpha) + \Delta_S^{-1}(s_{i+1}, \alpha)}{2}\right)$$

where (s_i, α) is the $\text{round}(\frac{n}{2})$ th largest element of $\langle S \rangle^n$.

3 Unbalanced Linguistic Sensory Evaluation Model

The aim of this contribution is to propose a new and adequate unbalanced linguistic scale to carry out the classification of olive oil, taking into account the nature of the uncertainty in sensory evaluation processes. The proposed scale will

be validated by using the linguistic sensory evaluation model proposed in [13] based on fuzzy linguistic approach. In this section, we point out general features of this model and describe its phases.

The unbalanced linguistic sensory evaluation model is a good option in the sensory evaluation process of olive oil because the reference ranges for classifying the olive oil are not symmetrical in the method proposed by IOC and this model offers a scale with different levels of discrimination on both sides to express the tasters' perceptions. Unlike the classical quantitative IOC method, the unbalanced linguistic sensory evaluation model does not need some statistical analysis because the fuzzy linguistic semantics manage the uncertainty involved in the tasters' perceptions. Despite this, it is noteworthy that the linguistic aggregation operator to compute the collective intensity for each sensory attribute and the reference ranges of intensities to classify the samples of olive oil are equivalent to the quantitative method proposed by IOC.

The linguistic sensory evaluation model with an unbalanced linguistic terms set consists of the following phases: evaluation framework, gathering sensory information, and rating samples [13]. These are described in the following subsections.

3.1 Evaluation Framework

It defines the structure of the sensory evaluation process: the set of tasters, the set of samples of olive oil that will be evaluated and, finally, the unbalanced linguistic scale in which tasters' perceptions will be expressed.

In order to define this scale, it is necessary to set its number of terms, its syntax, and its distribution. The semantic of each term is calculated with the algorithm proposed in [16] to build the semantics for an unbalanced linguistic terms set, using a linguistic hierarchy (LH) [23] and the linguistic 2-tuples representation model [19] (a detailed description about the algorithm can be found in [16]). So, the algorithm provides a *hierarchical semantic representation* $LH(S)$ for an unbalanced linguistic terms set $\mathcal{S} = \{s_i, \quad i = 0, \dots, g\}$ and obtains its representation in a LH .

Finally, in the evaluation framework, it is necessary to transform the reference ranges to classify the sample of olive oil into linguistic 2-tuples in the unbalanced linguistic scale.

3.2 Gathering Sensory Information

Once the framework has been defined to evaluate the set of samples of olive oil, the sensory information must be provided by the taster panel. In a profile sheet with the unbalanced linguistic scale fixed in the evaluation framework, each taster provides the intensity perceived about each sensory characteristic.

3.3 Rating Samples

This phase computes a collective intensity for each sensory attribute in order to classify each sample of olive oil, according to the perceived intensities. Due to the fact that the sensory evaluation model manages information expressed in an unbalanced linguistic scale, it is necessary to accomplish CW processes with this type of information. To do so, this linguistic sensory evaluation model uses the computational model for unbalanced linguistic term set presented in [16, 17] to compute the collective intensity for each sensory attribute. According to the collective intensity of the fruity attribute and the collective intensity of the defect perceived with the greatest intensity (negative attributes), as well as the reference ranges, each sample of olive oil is classified among one of four quality categories established: *virgin extra*, *virgin*, *ordinary*, and *lampante*.

4 New Unbalanced Linguistic Scale to Classify Olive Oil Samples

In this section, we present the analysis carried out with two taster panels of olive oil to propose an adequate unbalanced linguistic scale to classify a sample of olive oil. We then present the validation of the proposed scale, carrying out a sensory evaluation case study for a set of 30 samples of olive oil. Finally, we analyze and discuss the results.

4.1 New Proposed Unbalanced Linguistic Scale

In order to analyze and define an unbalanced linguistic scale, we selected two taster panels composed of 6 women and 10 men between 22 and 55 years old, being 16 tasters and 2 panel leaders. Both taster panels are accredited by OIC from 2008. The scale initially proposed (see Fig. 1a) in the unbalanced linguistic sensory evaluation model [13] is shown to the taster panels in order to analyze it and to propose a better alternative. After several meetings with panels, they agreed that 5 is an insufficient number of labels to measure tasters' perceptions in order to classify doubtful samples between two categories.

To overcome such a limitation, they propose an unbalanced linguistic scale with 7 linguistic terms. The distribution of the proposed scale is as follows: a central linguistic term, four terms on the left side, and two terms on the right side. The syntax provided for the panels and the semantic obtained for the algorithm to build the semantic for an unbalanced linguistic scale are illustrated in Fig. 1b.

Once the tasters had defined the unbalanced linguistic scale and their semantics were computed [16], it is necessary to transform the reference ranges proposed by

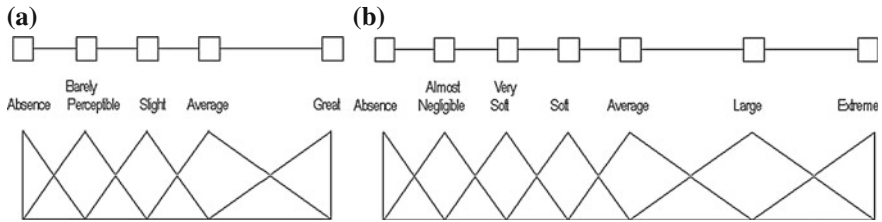


Fig. 1 a Initial linguistic scale b New proposed linguistic scale

IOC into linguistic 2-tuple on the unbalanced linguistic scale to classify the olive oil. The reference ranges expressed in linguistic 2-tuple to classify the olive oil are shown in the Table 1.

4.2 Procedure to Validate the Proposed Scale

In order to validate the proposed unbalanced linguistic scale, we conducted a sensory evaluation case study for 30 samples of olive oil, belonging to different categories, using the unbalanced linguistic sensory evaluation model, reviewed in Sect. 3.

The set of samples of olive oil were established with different profiles that included samples clearly pre-classified as one category and samples doubtful between two categories. The set of samples and their profiles were the following:

- 3 Extra Virgin (E1, E2, E3)
- 3 Extra Virgin (doubtful with Virgin) (EV1, EV2, EV3)
- 3 Virgin (doubtful with Extra Virgin) (VE1, VE2, VE3)
- 3 Virgin (V1, V2, V3)
- 3 Virgin (doubtful with Ordinary) (VO1, VO2, VO3)
- 3 Ordinary (doubtful with Virgin) (OV1, OV2, OV3)
- 3 Ordinary (O1, O2, O3)
- 3 Ordinary (doubtful with Lampante) (OL1, OL2, OL3)
- 3 Lampante (doubtful with Ordinary) (LO1, LO2, LO3)
- 3 Lampante (L1, L2, L3)

Table 1 Classification of olive oil

	Median defects (Med-n)	Median fruity (Med-f)
Extra virgin	Med-n=(absence, 0)	Med-f > (absence, 0)
Virgin	Med-n=(absence, 0)	Med-f=(absence, 0)
Virgin	(absence, 0) < Med-n ≤ (soft, 0)	
Ordinary	(soft, 0) < Med-n ≤ (average, 0.4)	
Lampante	(average, 0.4) < Med-n	

During the 2012 campaign of olive oil, the two panel leaders tasted different samples, of which 30 were selected according to the set of established profiles. Once the samples were selected, the sensory evaluation process began in order to validate the new proposed unbalanced linguistic scale. The sensory evaluation case study took place during seven weeks and was carried out following the test conditions and standards fixed by the method proposed by IOC.²

In the first week, two taster panels were trained in the unbalanced linguistic sensory evaluation model with the new proposed unbalanced linguistic scale shown in Fig. 1b.

During the next three weeks, each sample of olive oil was analyzed by a taster panel using the unbalanced linguistic sensory evaluation model. The set of 30 unidentified samples were distributed between the two taster panels. In a week, two tasting sessions were performed by each taster panel. Each taster panel analyzed among 2 and 4 samples of olive oil in each session.

In the last three weeks, in order to validate the classification obtained by the unbalanced linguistic sensory evaluation model, each taster panel tasted its samples by the IOC method. A comparative example of the sensory information and the classification obtained for the olive oil sample *EV1* with the unbalanced linguistic sensory evaluation model (linguistic information) and the IOC method (numerical information) are shown in Table 2.

4.3 Results and Discuss

In this section, we analyze the results of the conducted sensory evaluation case study.

The classification of olive oil samples was carried out by a panel leader, according to intensities of defects and the fruity attribute, following the guidelines of the unbalanced linguistic sensory evaluation model.

For each sample of the set, the category provided by the unbalanced linguistic sensory evaluation model, using the new proposed unbalanced linguistic scale, matched with the category computed by IOC method, based on the statistical analysis. The classifications obtained for the set of 30 olive oil samples, using both models, are shown in Table 3. Furthermore, it is noteworthy that each category also coincided with the pre-classification offered by the panel leader.

In view of the results, the set of tasters, who participated in the case study, pointed out that the high precision level required by the quantitative IOC sensory model is unnecessary to assess sensory attributes because this precision is not required to correctly classify olive oil samples. Furthermore, generally, this high precision level implies a long-term training of tasters and, sometimes, frustrated tasters.

² IOC/T.20/Doc. No 5 “Glass for oil tasting.”

Table 2 Sensory information about an evaluated sample of olive oil:EV1

Classification done by the panel leader: Extra Virgin (doubtful with Virgin)												
<i>Classification with the unbalanced linguistic sensory evaluation model: Extra Virgin</i>												
Taster	Fusty	Musty	Winey	Frostbitten	Rancid	Others	Fruity	Bitter	Pungent	Fusty	Musty	Winey
A	Alm. Neg.	Abs.	Abs.	Abs.	Abs.	Abs.	Alm. Neg.	Alm. Neg.	Alm. Neg.	Abs.	Abs.	Abs.
B	V. Soft	Abs.	Abs.	Abs.	Abs.	Abs.	V. Soft	Soft	Average	Abs.	Abs.	Abs.
C	Alm. Neg.	Abs.	Abs.	Abs.	Abs.	Abs.	V. Soft	Alm. Neg.	Alm. Neg.	Abs.	Abs.	Abs.
D	Abs.	Abs.	Abs.	Abs.	Abs.	Abs.	Soft	Soft	V. Soft	Abs.	Abs.	Abs.
E	Abs.	Abs.	Abs.	Abs.	Abs.	Abs.	Soft	V. Soft	V. Soft	Abs.	Abs.	Abs.
F	Abs.	Abs.	Abs.	Abs.	Abs.	Abs.	V. Soft	Soft	V. Soft	Abs.	Abs.	Abs.
G	Abs.	Abs.	Abs.	Abs.	Abs.	Abs.	V. Soft	V. Soft	Soft	Abs.	Abs.	Abs.
H	Abs.	Abs.	Abs.	Abs.	Abs.	Abs.	V. Soft	Alm. Neg.	Alm. Neg.	Abs.	Abs.	Abs.
Median	(Abs.,0)	(Abs.,0)	(Abs.,0)	(Abs.,0)	(Abs.,0)	(Abs.,0)	(V. Soft,0)	(V. Soft,0)	(V. Soft,0)	(Abs.,0)	(Abs.,0)	(Abs.,0)
<i>Classification with the IOC method: Extra Virgin</i>												
Taster	Fusty	Musty	Winey	Frostbitten	Rancid	Others	Fruity	Bitter	Pungent	Fusty	Musty	Winey
A	0	0	0	0	0	0	2,7	2,7	4	0	0	0
B	0	0	0	0	0	0	3,3	3	3,4	0	0	0
C	1,7	0	0	0	0	0	3,2	2,3	2,7	0	0	0
D	0	0	0	0	0	0	4,1	3,5	2,1	0	0	0
E	0	0	0	0	0	0	3,7	1,6	2,7	0	0	0
F	0	0	0	0	0	0	3	1,7	2,5	0	0	0
G	0	1,4	0	0	0	0	3	2,9	4,2	0	0	0
H	2	0	0	0	0	0	2	2,8	2,9	0	0	0
Median	0,00	0,00	0,00	0,00	0,00	0,00	3,10	2,75	2,80	0,00	0,00	0,00

Table 3 Classification obtained with both models

	Linguistic model	IOC method
Extra virgin	E1, E2, E3, EV1, EV2, EV3	E1, E2, E3, EV1, EV2, EV3
Virgin	VE1, VE2, VE3, V1, V2, V3, VO1, VO2, VO3,	VE1, VE2, VE3, V1, V2, V3, VO1, VO2, VO3,
Ordinary	OV1, OV2, OV3, O1, O2, O3, OL1, OL2, OL3	OV1, OV2, OV3, O1, O2, O3, OL1, OL2, OL3
Lampante	LO1, LO2, LO3, L1, L2, L3	LO1, LO2, LO3, L1, L2, L3

Therefore, the results of the sensory evaluation case study are very satisfactory. The proposed unbalanced linguistic scale offers more flexibility to express perceptions, models and manages consistently the uncertainty and vagueness presented in sensory evaluation process, and provides the same classification as the IOC method based on the statistical analysis and the opinion of a panel leader.

5 Conclusions

The sensory evaluation is a process in which the uncertainty and vagueness are present because the involved information is based on the knowledge acquired via human senses. The use of linguistic information to model and manage such uncertainty is a suitable framework. One part of the success of a sensory evaluation process depends on the correct definition of the linguistic scale used to measure the sensory information. In this contribution, we have analyzed and proposed with two taster panels an unbalanced linguistic scale to classify a sample of olive oil in one category of the four categories of quality established by the IOC. The proposed unbalanced linguistic scale has been validated in a sensory evaluation case study, using an unbalanced linguistic sensory evaluation model based on the fuzzy linguistic approach, proposed in our previous research [13]. The main result of this case study is that the proposed unbalanced linguistic scale provides more flexibility to express the perceptions, offering the same classification as the method proposed by IOC based on statistical analysis.

Acknowledgments This contribution has been supported by the research project AGR-6487.

References

1. Dijksterhuis G (1997) *Multivariate data analysis in sensory and consumer science*. Food and Nutrition (Press Inc.), Trumbull, Connecticut, USA
2. Ruan D, Zeng X (eds) (2004) *Intelligent sensory evaluation: methodologies and applications*. Springer, New York
3. Martínez L (2007) Sensory evaluation based on linguistic decision analysis. *Int. J. Approximate Reasoning* 44(2):148–164
4. Zadeh L (1975) The concept of a linguistic variable and its applications to approximate reasoning. *Inform Sci, Part I, II, III, vol 8, 9*, pp 199–249, 301–357, 43–80
5. Zadeh L (1965) Fuzzy sets. *Inform Control* 8(3):338–353
6. Chen Y, Zeng X, Happiette M, Bruniaux P, Ng R, Yu W (2009) Optimisation of garment design using fuzzy logic and sensory evaluation techniques. *Eng Appl Artif Intell* 22(2):272–282
7. Perrot N, Ioannou I, Allais I, Curt C, Hossenlopp J, Trystram G (2006) Fuzzy concepts applied to food product quality control: a review. *Fuzzy Sets Syst* 157(9):1145–1154
8. Jaya S, Das H (2003) Sensory evaluation of mango drinks using fuzzy logic. *J Sens Stud* 18(2):163–176
9. Siniya V, Mishra H (2011) Fuzzy analysis of sensory data for quality evaluation and ranking of instant green tea powder and granules. *Food Bioprocess Technol* 4(3):408–416
10. Russo L, Albanese D, Siettos C, di Matteo M, Crescitelli S (2012) A neuro-fuzzy computational approach for multicriteria optimisation of the quality of espresso coffee by pod based on the extraction time, temperature and blend. *Int J Food Sci Technol* 47(4):837–846
11. Lee S, Kwon YA (2007) Study on fuzzy reasoning application for sensory evaluation of sausages. *Food Control* 18(7):811–816
12. Routray W, Mishra H (2012) Sensory evaluation of different drinks formulated from dahi (indian yogurt) powder using fuzzy logic. *J Food Processing Preserv* 36(1):1–10
13. Martínez L, Espinilla M, Liu J, Pérez L, Sánchez P (2009) An evaluation model with unbalanced linguistic information applied to olive oil sensory evaluation. *J Multiple Valued Logic Soft Comput* 15(2–3):229–251
14. Blery E, Sfetsiou E (2008) Marketing olive oil in Greece. *Br Food J* 110(11):1150–1162
15. de Graaff J, Duran Zuazo V-H, Jones N, Fleskens L (2008) Olive production systems on sloping land: prospects and scenarios. *J Environ Manage* 89(2):129–139
16. Herrera F, Herrera-Viedma E, Martínez L (2008) A fuzzy linguistic methodology to deal with unbalanced linguistic term sets. *IEEE Trans Fuzzy Syst* 16(2):354–370
17. Martínez L, Herrera F (2012) An overview on the 2-tuple linguistic model for computing with words in decision making: Extensions, applications and challenges. *Inf Sci* 207(1):1–18
18. Martínez L, Ruan D, Herrera F (2010) Computing with words in decision support systems: an overview on models and applications. *Int J Comput Intell Syst* 3(4):382–395
19. Herrera F, Martínez L (2000) A 2-tuple fuzzy linguistic representation model for computing with words. *IEEE Trans Fuzzy Syst* 8(6):746–752
20. Wei G (2011) Some harmonic aggregation operators with 2-tuple linguistic assessment information and their application to multiple attribute group decision making. *Int J Uncertainty Fuzziness Knowl Based Syst* 19(6):977–998
21. Xu Y, Wang H (2011) Approaches based on 2-tuple linguistic power aggregation operators for multiple attribute group decision making under linguistic environment. *Appl Soft Comput J* 11(5):3988–3997
22. Yang W, Chen Z (2012) New aggregation operators based on the choquet integral and 2-tuple linguistic information. *Expert Syst Appl* 39(3):2662–2668
23. Herrera F, Martínez L (2001) A model based on linguistic 2-tuples for dealing with multigranularity hierarchical linguistic contexts in multiexpert decision-making. *IEEE Trans Syst Man Cybern B Cybern* 31(2):227–234

Algorithm to Find Ground Instances in Linguistic Truth-Valued Lattice-Valued First-Order Logic $\mathcal{L}_{V(n \times 2)}F(X)$

Xiaomei Zhong, Yang Xu and Peng Xu

Abstract α -Resolution-based automated reasoning in linguistic truth-valued lattice-valued first-order logic $\mathcal{L}_{V(n \times 2)}F(X)$ based on linguistic truth-valued lattice implication algebra $\mathcal{L}_{V(n \times 2)}$ can be equivalently transformed into that in lattice-valued propositional logic $\mathcal{L}_n P(X)$, and during this equivalent transformation, the key step is to find out the corresponding ground instance. To solve the above-mentioned problem, this paper gives an algorithm to find the corresponding ground instance of generalized-clause sets satisfying certain conditions in $\mathcal{L}_{V(n \times 2)}F(X)$, which lays the foundation for the application of α -resolution-based automated reasoning in $\mathcal{L}_{V(n \times 2)}F(X)$, to some extent.

Keywords Algorithm to find ground instances · Resolution-based automated reasoning · Linguistic truth-valued lattice-valued first-order logic · Lattice implication algebra

1 Introduction

Since resolution principle based on classical logic was established by Robinson [1] in 1965, resolution-based automated reasoning has been widely applied in some areas, such as theorem automated proving, software verification, and problem solving. As classical logic can only deal with certain questions, in order to deal with fuzziness and incomparability, Xu et al., established a kind of multi-valued logic system—lattice-valued logic system based on lattice implication algebra.

X. Zhong (✉) · Y. Xu · P. Xu
School of Mathematics, Southwest Jiaotong University, Chengdu 610031, Sichuan,
People's Republic of China
e-mail: zhongxm@126.com

To research resolution-based automated reasoning in the above lattice-valued logic, in 2000 and 2001, Xu et al. [2, 3] proposed a kind of gradational resolution principle— α -resolution principle in lattice-valued propositional logic LP(X)-based lattice implication algebra and the corresponding first-order logic LF(X), respectively, and further proved its soundness and weak completeness theorems. In order to expand the applicable range of α -resolution-based automated reasoning, in 2010, Xu et al. [4] extended α -resolution principle into a more general form—the general form of α -resolution principle in LP(X) and LF(X), respectively—and established its soundness and completeness simultaneously, which lays a good theoretical basis for practical application of automated reasoning based on the general form of α -resolution principle.

In real life, people do the reasoning, decision-making, and judgment usually by natural language. Hence, it is very necessary to use linguistic truth values to implement α -resolution-based automated reasoning in lattice-valued logic based on lattice implication algebra. Based on this idea, in 2006, Xu et al. [5] established a linguistic truth-valued lattice implication algebra $\mathcal{L}_{V(n \times 2)}$ by some linguistic truth values, which are frequently used in daily life. In 2012 [6], we equivalently transformed the general form of α -resolution principle in linguistic truth-valued lattice-valued first-order logic $\mathcal{L}_{V(n \times 2)} F(X)$ into that in lattice-valued propositional logic $\mathcal{L}_n P(X)$ based on Łukasiewicz implication algebra \mathcal{L}_n . During this equivalent transformation, the key step is to find out the corresponding ground substitution, that is, the corresponding ground instance. Therefore, to further apply α -resolution-based automated reasoning in $\mathcal{L}_{V(n \times 2)} F(X)$ to the actual problem, it is very important for us to find the corresponding ground instance of generalized-clause sets. In this paper, we will give an algorithm to find the corresponding ground instance of generalized-clause sets satisfying certain conditions in $\mathcal{L}_{V(n \times 2)} F(X)$.

This paper is organized as follows: in Sect. 2, some preliminary relevant concepts about lattice-valued first-order logic LF(X) are reviewed; In Sect. 3, the algorithm to find ground instances of generalized-clause sets satisfying certain conditions is designed.

2 Preliminaries

In the following, we will review some elementary concepts and conclusions of lattice-valued first-order logic LF(X) based on lattice implication algebra. We refer the readers to [7] for more details.

Definition 2.1 Xu et al. [7] (*Łukasiewicz implication algebra on finite chain*) Let $L_n = \{a_i \mid i = 1, 2, \dots, n\}$, $a_1 < a_2 < \dots < a_n$. For any $1 \leq j, k \leq n$, define

$$\begin{aligned} a_j \vee a_k &= a_{\max\{j,k\}}, \quad a_j \wedge a_k = a_{\min\{j,k\}}, \quad (a_j)' = a_{n-j+1}, \\ a_j \rightarrow a_k &= a_{\min\{n-j+k, n\}}. \end{aligned}$$

Then $(L_n, \vee, \wedge, \iota, \rightarrow, a_1, a_n)$ is a LIA, denoted as \mathcal{L}_n .

Definition 2.2 Xu et al. [7] Suppose V and F are the set of variable symbols and that of functional symbols in lattice-valued first-order $LF(X)$, respectively, the set of terms of $LF(X)$ is defined as the smallest set \mathcal{F} satisfying the following conditions:

- (1) $V \subseteq \mathcal{F}$,
- (2) For any $n \in \mathbb{N}$. if $f^{(n)} \in F$, then for any $t_0, t_1, \dots, t_n \in \mathcal{F}$, $f^{(n)}(t_0, t_1, \dots, t_n) \in \mathcal{F}$.

Remark 2.1 $f^{(0)}$ is .

Definition 2.3 Xu et al. [7] Suppose P is the predicate symbol set in $LF(X)$. The set of atoms of $LF(X)$ is defined as the smallest \mathcal{A}_t satisfying the following condition: For any $n \in \mathbb{N}$, if $P^{(n)} \in P$, then $P^{(n)}(t_0, t_1, \dots, t_n) \in \mathcal{A}_t$ for any $t_0, t_1, \dots, t_n \in \mathcal{F}$.

Remark 2.2 $P^{(0)}$ is specified as a certain element in L .

Definition 2.4 Xu et al. [7] The set of formulas of $LF(X)$ is defined as the smallest set \mathcal{F} satisfying the following conditions:

- (1) $\mathcal{A}_t \subset \mathcal{F}$,
- (2) If $p, q \in \mathcal{F}$, then $p \rightarrow q \in \mathcal{F}$,
- (3) If $p \in \mathcal{F}$, x is a free variable in p , then $(\forall x) p, (\exists x) p \in \mathcal{F}$.

Remark 2.3 In fact, if $p, q \in \mathcal{F}$, then $p', p \vee q, p \wedge q, p \leftrightarrow q \in \mathcal{F}$.

Definition 2.5 Xu et al. [7] Suppose $G \in \mathcal{F}$, F_G is the set of all functional symbols occurring in G , P_G is the set of all predicate symbols occurring in G , and $D (\neq \phi)$ is the domain of interpretation. An interpretation of G over D is a triple $I_D = \langle D, \mu_D, V_D \rangle$, where

$$\begin{aligned} \mu_D : F_G \rightarrow U_D &= \left\{ f_D^{(n)} : D^n \rightarrow D \mid n \in \mathbb{N} \right\} \\ f^{(0)} \mapsto f_D^{(0)}, f_D^{(0)}(D^0) &= \left\{ f_D^{(0)} \right\} \subseteq D, D^{(0)} \text{ is a non - empty set} \\ f^{(n)} \mapsto f_D^{(n)}(n \in \mathbb{N}^+), \end{aligned}$$

$$\begin{aligned} V_D : P_G \rightarrow V_D &= \left\{ P_D^{(n)} : D^n \rightarrow L \mid n \in \mathbb{N} \right\} \\ p^{(0)} \mapsto p_D^{(0)}, p_D^{(0)}(D^0) &= \left\{ p_D^{(0)} \right\} \subseteq L \\ p^{(n)} \mapsto p_D^{(n)}(n \in \mathbb{N}^+). \end{aligned}$$

Definition 2.6 Xu et al. [7] A lattice-valued first-order logical formula G in lattice-valued first-order logic system $\text{LF}(X)$ is called an extremely simple form, in short ESF, if a lattice-valued first-order logical formula G^* obtained by deleting any literal or implication term occurring in G is not equivalent to G .

Definition 2.7 Xu et al. [7] A lattice-valued first-order logical formula G in lattice-valued first-order logic system $\text{LF}(X)$ is called an indecomposable extremely simple form, in short IESF, if the following two conditions hold:

- (1) G is an ESF containing connective \rightarrow and $'$ at most,
- (2) For any $H \in \mathcal{F}$, if $H \in \overline{G}$ in $\overline{\text{LF}(X)}$, then H is an ESF containing connectives \rightarrow and $'$ at most, where $\overline{\text{LF}(X)} = (\overline{\mathcal{F}}, \vee, \wedge, ', \rightarrow)$ is a lattice implication algebra and $\overline{\mathcal{F}}$ is the set composed of all equivalence class of logical formulae in $\text{LF}(X)$.

All the literals and IESFs in $\text{LF}(X)$ are called generalized literals. The disjunction of a finite number of generalized literals is a generalized-clause.

In the following, generalized-clauses always belong to a generalized-Skölem standard form, that is, for any generalized-clause C , all variables of C are bound variables with the quantifier \forall , and generalized-clauses C_1, C_2, \dots, C_m ($m \geq 3$) have no common variables. In addition, the definitions of atom, ground atom, ground substitution, and ground instance are the same as those in classical logic.

Theorem 2.1 Zhong [8] Let $C = C_1 \wedge C_2 \wedge \dots \wedge C_m$, where C_1, C_2, \dots, C_m are generalized-clauses in lattice-valued first-order logic $\mathcal{L}_n F(X)$ based on Łukasiewicz implication algebra \mathcal{L}_n , and $\alpha \in \mathcal{L}_n$. Then, $C \leq \alpha$ if and only if there exists a ground substitution θ such that $C_1^\theta \wedge C_2^\theta \wedge \dots \wedge C_m^\theta \leq \alpha$, that is, $C^\theta \leq \alpha$.

3 Algorithm for Finding Ground Instances

In Ref. [6], we obtained the conclusion that the general form of α -resolution principle in linguistic truth-valued lattice-valued first-order logic $\mathcal{L}_{V(n \times 2)} F(X)$ can be equivalently transformed into that in lattice-valued propositional logic $\mathcal{L}_{V_n} P(X)$ based on Łukasiewicz implication algebra \mathcal{L}_{V_n} with linguistic truth values as its elements. In fact, firstly, the general form of α -resolution principle in $\mathcal{L}_{V(n \times 2)} F(X)$ is equivalently transformed into that in lattice-valued first-order logic $\mathcal{L}_{V_n} F(X)$ based on \mathcal{L}_{V_n} , and then the general form of α -resolution principle in $\mathcal{L}_{V_n} F(X)$ is further equivalently transformed into that in $\mathcal{L}_{V_n} P(X)$. To this equivalent transformation, the most critical step is later, that is, finding out the corresponding ground instance of generalized-clause sets in $\mathcal{L}_{V_n} F(X)$. Although the existence of the corresponding ground instance in $\mathcal{L}_{V_n} P(X)$, was proved in Ref. [6] but a specific method of looking for the ground did not be proposed instance. As lattice implication algebra \mathcal{L}_{V_n} has the same structure as that of Łukasiewicz implication algebra \mathcal{L}_n , without loss of generality, we will give an

algorithm to look for the ground instance of a class of logical formulae in lattice-valued first-order logic $\mathcal{L}_n F(X)$ in the following.

Before giving the specific algorithm, we need to point out the conditions satisfied by logical formula F firstly. So, we have the following condition.

Condition 3.1 Let F be a logical formula in lattice-valued first-order logic $\mathcal{L}_n F(X)$, P an atom of F , W_P the sequence composed of all the atoms, which occur in F and have the same predicate symbol as that of P . Suppose D_k is the sequence composed of the k th component of all the atoms in W_P , where $k \in N^+$. For any D_k , D_k satisfies the following conditions:

- (1) There is a constant or function symbol occurring in D_k at most, and constants and function symbols do not occur in D_k simultaneously.
- (2) For any function symbol f occurring in D_k , f satisfies the following conditions:
 1. for any variable x in D_k , x is not a independent variable of f .
 2. let $H_k = \{(x_{1j}, x_{2j}, \dots, x_{rj}) \mid x_{ij}$ is the j th component of the independent variable of f_i ($i = 1, 2, \dots, r$), $j \in N^+$, f_1, f_2, \dots, f_r is the sequence composed of all function symbols occurring in $D_k\}$. For any $(x_{1j}, x_{2j}, \dots, x_{rj}) \in H_k$, there is a constant occurring in $x_{1j}, x_{2j}, \dots, x_{rj}$ at most.

Suppose logical formula F satisfying Condition 3.1, the algorithm is as follows:

Step 1: Let $k = 1$, $W_P^k = W_P$.

Step 2: If for any atom E in W_P^k , E does not include variables, then the algorithm stops and W_P^k is the sequence composed of the ground atoms of all atoms in W_P . Otherwise, go to Step 3.

Step 3: Find out the component sequence D_k of W_P^k denoted as $D_k = D_{k1}, D_{k2}, D_{k3}$, where D_{k1} is the sequence composed of all constants in D_k , D_{k2} is the sequence composed of all variables in D_k , D_{k3} is the sequence composed of all function symbols in D_k .

Step 4: If D_{k1} does not exist, then go to Step 5. Otherwise, go to Step 7.

Step 5:

1. If D_{k2} and D_{k3} both exist, then replace all the variables in D_{k2} with the function in D_{k3} , and denote the obtained sequence as D_{k2}^* , go to Step 6.
2. If D_{k2} exists and D_{k3} does not exist, then replace all the variables in D_{k2} with the same constant and go to Step 8.
3. If D_{k2} does not exist and D_{k3} exists, then let D_{k2}^* be an empty sequence and go to Step 6.

Step 6: Let the sequence composed of D_{k2}^* and D_{k3} is f_1, f_2, \dots, f_w , $w \in N^+$, $V_f = \{(x_{1j}, x_{2j}, \dots, x_{wj}) \mid x_{lj}$ is the j th component of function f_l , $l = 1, 2, \dots, w$, $j \in N^+\}$. For any $(x_{1j}, x_{2j}, \dots, x_{wj}) \in V_f$: if $x_{1j}, x_{2j}, \dots, x_{wj}$ are variables, then replace $x_{1j}, x_{2j}, \dots, x_{wj}$ with the same constant; if there exist a constant in $x_{1j}, x_{2j}, \dots, x_{wj}$, then replace all the variables in $x_{1j}, x_{2j}, \dots, x_{wj}$ as the constant occurring in $x_{1j}, x_{2j}, \dots, x_{wj}$. For any variable and function in D_{k2}, D_{k3} , do the corresponding modification and go to Step 8.

Step 7:

1. If D_{k2} exists, then replace all variables in D_{k2} with the constant in D_{k1} and go to Step 8.
2. If D_{k2} does not exist, then go to Step 8.

Step 8: For any atom in W_P^k , replace the k th component of E with the corresponding element in D_k and denote the obtained sequence as W_P^{k+1} . Let $k = k+1$ and go to Step 2.

The flow chart of this algorithm is shown in the following figure, that is, Fig. 1.

For any atom sequence W_Q of F , we can obtain a ground sequence W_Q^k composed of the ground atoms of all atoms in W_Q by the above algorithm. After replacing all the atoms Q of F with the corresponding ground atoms in W_Q^k we can get a ground instance F^* of F .

Theorem 3.1 *Suppose W is a sequence composed of some atoms with the same predicate symbol in lattice-valued first-order logic $\mathcal{L}_n F(X)$, and W satisfies Condition 3.1. Then, W^k is a sequence composed of ground atoms of all atoms in W if and only of the algorithm stops at Step 2.*

Proof The conclusion holds obviously.

Example 3.1 Suppose $C_1 = (M(f(a, x_1)) \rightarrow N(x_2))' \vee (M(f(x_3, b)) \rightarrow N(a))'$, $C_2 = (M(f(a, y_1)) \rightarrow N(a)) \vee (M(f(y_2, b)) \rightarrow N(y_3))$, $C_3 = (N(a) \rightarrow P(a, z_1)) \vee P(z_2, h(a, z_3))$ are generalized-clauses in lattice-valued first-order logic $\mathcal{L}_n F(X)$, denote as $S = C_1 \wedge C_2 \wedge C_3$, where $x_1, x_2, x_3, y_1, y_2, y_3, z_1, z_2, z_3$ are variables and a, b are constants.

Since there are three atom sequences with the same predicate symbol as follows:

$$\begin{aligned} W_1 &= M(f(a, x_1)), M(f(x_3, b)), M(f(a, y_1)), M(f(y_2, b)), \\ W_2 &= N(x_2), N(a), N(a), N(y_3), N(a), \\ W_3 &= P(a, z_1), P(z_2, h(a, z_3)). \end{aligned}$$

According to the above algorithm, we can obtain the following ground atom set:

For W_1 : $W_1^1 = W_1$, $D_1 = f(a, x_1), f(x_3, b), f(a, y_1), f(y_2, b)$, that is, D_{11}, D_{12} do not exist and D_{13} exists. According to Step 5 and Step 6 of the algorithm, we have $D_1 = f(a, b), f(a, b), f(a, b), f(a, b)$. Hence, $W_1^2 = M(f(a, b)), M(f(a, b)), M(f(a, b)), M(f(a, b))$ is the sequence composed of the ground atoms of all atoms in W_1 .

For W_2 : $W_2^1 = W_2$, $D_1 = x_2, a, a, y_3, a$, that is, $D_{11} = a, a, a$, $D_{12} = x_2, y_3$ and D_{13} does not exist. According to Step 7 of the algorithm, we have $D_1 = a, a, a, a, a$. Hence, $W_2^2 = N(a), N(a), N(a), N(a), N(a)$ is the sequence composed of the ground atoms of all atoms in W_2 .

For W_3 : $W_3^1 = W_3$, $D_1 = a, z_2$, that is, $D_{11} = a$, $D_{12} = z_2$ and D_{13} does not exist. According to Step 7 of the algorithm, we have $D_1 = a, a$. Hence, $W_3^2 = P(a, z_1), P(a, h(a, z_3))$. Since W_3^2 is not the sequence composed of the ground atoms of all atoms in W_3 , we have $D_2 = z_1, h(a, z_3)$, that is, D_{21} does not exist, $D_{22} = z_1$ and

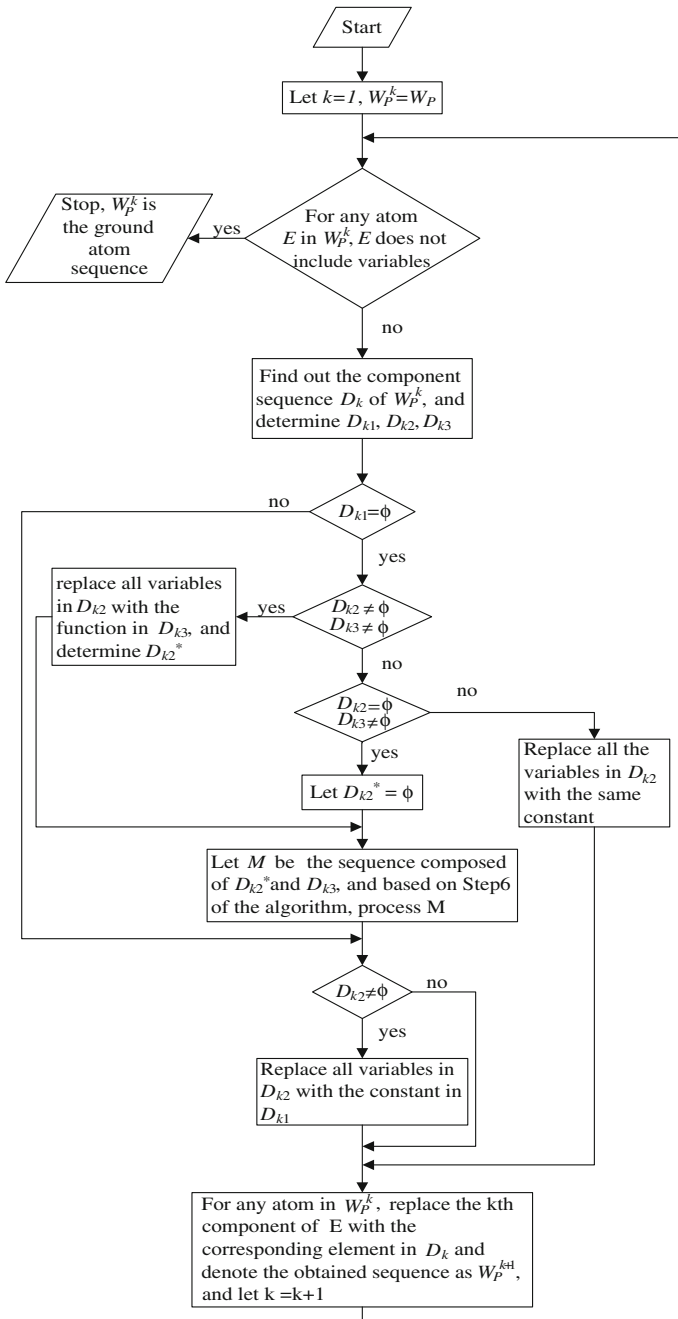


Fig. 1 Flow chart of the algorithm

$D_{23} = h(a, z_3)$. According to Step 5 and Step 6 of the algorithm, we have $D_{22} = h(a, c)$, $D_{23} = h(a, c)$, where, c is a constant of $\mathcal{L}F(X)$. So, $W_3^3 = P(a, h(a, c))$, $P(a, h(a, c))$ is the sequence composed of the ground atoms of all atoms in W_3 .

Therefore, $C_1^* = (Mf(a, b) \rightarrow N(a))'$, $C_2^* = (Mf(a, b) \rightarrow N(a))$, $C_3^* = (N(a) \rightarrow P(a, h(a, c))) \vee P(a, h(a, c))$, $S^* = C_1^* \wedge C_2^* \wedge C_3^*$ is a ground instance of S .

Suppose S is a generalized-clause set in lattice-valued first-order logic $\mathcal{L}_n F(X)$. If S satisfies Condition 3.1, then we can obtain a ground instance of all atoms in S by the above algorithm and further get a ground instance of S .

Theorem 3.2 *Suppose F is a logical formula in lattice-valued first-order logic $\mathcal{L}_n F(X)$, and F satisfies Condition 3.1. F^* is a ground instance of F obtained by the above algorithm and $\alpha \in L_n$. Then $F \leq \alpha$ if and only if $F^* \leq \alpha$.*

Proof (\Rightarrow) Suppose W is the sequence composed of atoms with the same predicate symbol in F , and W^* is the sequence composed of the corresponding ground atoms in F^* of all atoms in W . Since F^* is a ground instance of F obtained by the above algorithm, we have the result that all elements of W^* are the same. For any ground substitution θ of F , suppose W^θ is the sequence obtained by doing substitution θ on all atoms in W . As $F \leq \alpha$, according to Theorem 2.1, there exists ground substitution θ° such that $F^{\theta^\circ} \leq \alpha$. If $F^{\theta^\circ} = F^*$, then the conclusion holds obviously. If $F^{\theta^\circ} \neq F^*$, then let V^{θ° be the truth-value set of F^{θ° in $\mathcal{L}_n F(X)$, that is, $V^{\theta^\circ} = \{l | I_D = \langle D, \mu_D, \nu_D \rangle \text{ is an interpretation of } \mathcal{L}_n F(X), \nu_D(F^{\theta^\circ}) = l\}$. Since W^θ may include different elements, we can obtain $V^* \subseteq V^{\theta^\circ}$, where V^* is the truth-value set of F^* in $\mathcal{L}_n F(X)$. As for any $l \in V^{\theta^\circ}$, $l \leq \alpha$, we have $F^* \leq \alpha$.

(\Leftarrow) Since F^* is a ground instance of F and $F^* \leq \alpha$, we have $F \leq \alpha$ obviously.

Since a generalized-clause set is also a logical formula, according to Theorem 3.2, we can obtain similar result about generalized-clause sets in lattice-valued first-order logic $\mathcal{L}_n F(X)$.

Example 3.2 Suppose $C_1 = P(x_1, a) \rightarrow Q(x_2)$, $C_2 = (P(b, y_1) \rightarrow R(y_2))' \vee (P(y_3, a) \rightarrow Q(y_4))'$, $C_3 = (Q(a) \rightarrow R(z_1))' \vee Q(z_2)'$ are generalized-clauses in lattice-valued first-order logic $\mathcal{L}_9 F(X)$ based on Łukasiewicz implication algebra with nine elements, denote as $S = C_1 \wedge C_2 \wedge C_3$, where $x_1, x_2, y_1, y_2, y_3, y_4, z_1, z_2$ are variables, a, b are constants and $\alpha = a_6$. Then $S \leq \alpha$.

There are three atom sequences with the same predicate symbol as follows:

$$W_1 = P(x_1, a), P(b, y_1), P(y_3, a),$$

$$W_2 = Q(x_2), Q(y_4), Q(a), Q(z_2)',$$

$$W_3 = R(y_2), R(z_1).$$

According to the above algorithm, we can obtain the following ground atom set:

- (1) $W_1^2 = P(b, a), P(b, a), P(b, a)$ is the sequence composed of the ground atoms of all atoms in W_1 .

- (2) $W_2^1 = Q(a), Q(a), Q(a), Q(a)'$ is the sequence composed of the ground atoms of all atoms in W_2 .
- (3) $W_3^1 = R(a), R(a)$ is the sequence composed of the ground atoms of all atoms in W_3 .

Therefore, $C_1^* = P(b, a) \rightarrow Q(a)$, $C_2^* = (P(b, a) \rightarrow R(a))' \vee (P(b, a) \rightarrow Q(a))'$, $C_3^* = (Q(a) \rightarrow R(a))' \vee Q(a)'$, and $S^* = C_1^* \wedge C_2^* \wedge C_3^*$ is a ground instance of S .

As $S^* = C_1^* \wedge C_2^* \wedge C_3^*$, we have $S^* = ((P(b, a) \rightarrow Q(a)) \wedge (P(b, a) \rightarrow R(a))' \wedge (Q(a) \rightarrow R(a))') \vee ((P(b, a) \rightarrow Q(a)) \wedge (P(b, a) \rightarrow R(a))' \wedge Q(a)') \vee ((P(b, a) \rightarrow Q(a)) \wedge (P(b, a) \rightarrow Q(a))' \wedge (Q(a) \rightarrow R(a))') \vee ((P(b, a) \rightarrow Q(a)) \wedge (P(b, a) \rightarrow Q(a))' \wedge Q(a)')$.

Since $(P(b, a) \rightarrow Q(a)) \wedge (P(b, a) \rightarrow R(a))' \wedge (Q(a) \rightarrow R(a))' \leq a_6$, $(P(b, a) \rightarrow Q(a)) \wedge (P(b, a) \rightarrow R(a))' \wedge Q(a)' \leq a_6$ and $(P(b, a) \rightarrow Q(a)) \wedge (P(b, a) \rightarrow Q(a))' \leq a_6$, we have $S^* \leq a_6$.

Hence, according to Theorem 3.2, we can obtain $S \leq a_6$.

4 Conclusion

In this paper, we give an algorithm to find ground instances of logical formulae. Concretely, for a class of logical formula F (i.e., F satisfies Condition 3.1) in lattice-valued first-order logic $\mathcal{L}_n F(X)$, we can obtain a ground instance F^* of F by this algorithm, and the unsatisfiability of F^* and F is consistent. Since F^* is a logical formula in lattice-valued propositional logic $\mathcal{L}_n P(X)$, the determination of unsatisfiability of F becomes relatively simple.

Acknowledgments This work is supported by National Science Foundation of China (Grant No. 61175055, 61100046), Sichuan Key Technology Research and Development Program (Grant No. 2011FZ0051), Radio Administration Bureau of MIIT of China(Grant No. [2011]146), China Institution of Communications(Grant No. [2011]051).

References

1. Robinson JP (1965) A machine-oriented logic based on the resolution principle. J ACM 12:23–41
2. Xu Y, Ruan D, Kerre EE, Liu J (2000) α -Resolution principle based on lattice-valued propositional logic LP(X). Inform Sci 130:195–223
3. Xu Y, Ruan D, Kerre EE, Liu J (2001) α -Resolution principle based on lattice-valued first-order lattice-valued logic LF(X). Inf Sci 132:221–239
4. Xu Y, Zhong XM, Liu J, Chen SW (2010) General form of α -resolution principle based on lattice-valued logic with truth-value in lattice implication algebras, Information Sciences, under review
5. Xu Y, Chen SW, Ma J (2006) Linguistic truth-valued lattice implication algebra and its properties. In: Proceedings of IMACS multi conference on computational engineering in systems applications (CESA2006), Beijing, China, pp 1413–1418

6. Zhong XM, Xu Y, Liu J, Ruan D, Chen SW (2012) General form of α -resolution principle for linguistic truth-valued lattice-valued logic. *Soft Comput* 16:1767–1781
7. Xu Y, Ruan D, Qin KQ, Liu J (2003) *Lattice-valued logic—An alternative approach to treat fuzziness and incomparability*. Springer, Berlin
8. Zhong XM (2012) *Study on α -quasi-lock semantic resolution automated reasoning based on lattice-valued logic*, Doctor Degree Dissertation, Southwest Jiaotong University

Estimating Online Review Helpfulness with Probabilistic Distribution and Confidence

Zunqiang Zhang, Qiang Wei and Guoqing Chen

Abstract Product review helpfulness information is useful knowledge for consumers in their online shopping decision processes. Unlike the traditional method using the simple voting percentages, this paper proposes a new method for estimating the degrees of helpfulness with two features. One is to take into account the helpfulness distribution information on all reviews of concern in determination of helpfulness degrees; the other is to construct confidence intervals (CIs) of helpfulness to distinguish different reviews with the same voting percentage. Both synthetic and real data experiments, along with an illustrative example, reveal that the proposed method is superior to the traditional one in light of estimation accuracy.

Keywords Online reviews · Review helpfulness · Review recommendation · Helpfulness estimation

1 Introduction

With the advent of Web 2.0, online users show great passion to post reviews to share their experiences/opinions on products/services. Since potential buyers could make better purchase decisions by browsing online reviews to learn knowledge about items from other consumers [1, 2], the services such as recommendation aids provided by Web sites based upon the reviews are considered helpful for

Z. Zhang · Q. Wei (✉) · G. Chen
School of Economics and Management, Tsinghua University, Beijing 100084, China
e-mail: weiq@sem.tsinghua.edu.cn

Z. Zhang
e-mail: zhangzq2.09@sem.tsinghua.edu.cn

G. Chen
e-mail: chengq@sem.tsinghua.edu.cn

generating additional sales and revenues [3]. However, due to the fact that the amount of reviews may be quite large with various levels of quality, information overload becomes a challenging problem [1–3]. To cope with the problem and facilitate users acquiring knowledge from other consumers, many e-commerce services/Web sites use a so-called review helpfulness voting mechanism to collect users' opinion about reviews' helpfulness and rank reviews based on the votes. For instance, Amazon.com would display a review to a potential buyer with previous buyers' voting results of its helpfulness, for example, "137 of 178 people found the review helpful," as illustrated in Fig. 1.

Though the helpfulness voting mechanism is regarded useful, it suffers from some biases where most online reviews receive very few helpfulness votes even in well-known Web sites, such as Amazon [4], Youtube [5], and IMDB [6], giving rise to an increasing number of research and application attempts at review helpfulness prediction [4–15].

For review helpfulness prediction, one of the core steps is to estimate the helpfulness degrees of voted reviews to train prediction models [4–11, 13–15]. A common and straight method is to calculate the percentage of helpful votes (e.g., the helpfulness degree is 137/178 for previous example) [6, 9, 12, 16], which is also widely adopted by many web services. Although this way (hereafter referred to as the traditional method) is direct and somewhat intuitive, the ranking results based on these estimated helpfulness degrees have some limitations [4, 12], which need to be further improved.

First, the traditional method ignores the helpfulness distribution information on all reviews of concern. Due to the large amount of helpfulness voting information accumulated in the web, the helpfulness distribution of reviews with similar characteristics may be utilized to enhance the estimation of helpfulness for a particular review. Depending upon the application contexts, the distributions could be built accordingly, such as for the same platform, community, product type, or product item. Second, the traditional method does not reflect different levels of confidence on estimated helpfulness. In other words, all reviews with the same percentage of helpful votes (e.g., 9/10 vs. 90/100) are treated equally (in terms of the level of confidence), which may not be intuitively appealing. For example, with helpfulness degree 0.9, "9 of 10 people found the review helpful" may

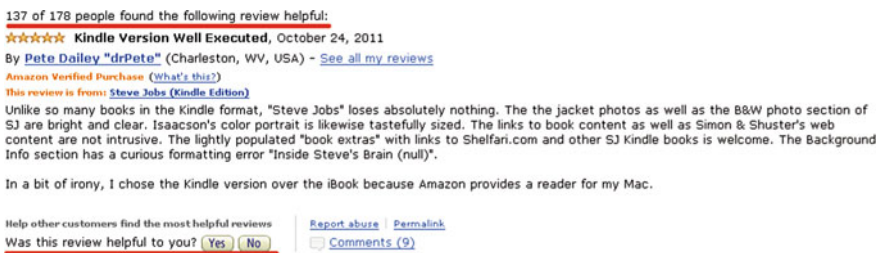


Fig. 1 Example of Amazon's review helpfulness voting mechanism (from <http://www.amazon.com/Steve-Jobs-ebook/product-reviews/B004W2UBYW/>)

provide the users with a lower level of confidence than “90 of 100 people found the review helpful.”

To remedy such limitations, this paper is aimed to propose a new estimation method incorporating helpfulness distribution information and providing confidence on estimated results. Section 2 introduces the theoretical explanation. Section 3 provides an illustrative example. The experiments on synthetic and real data are presented in Sect. 4. The conclusion is presented in Sect. 5.

2 A New Method for Helpfulness Estimation

For an online review, the traditional method uses a percentage of helpful votes (x) out of total votes (n) as the review’s estimated helpfulness degree x/n . It could be regarded as an exact helpfulness degree for a review if all the people concerned (i.e., the population) had voted this review, which is generally impossible. In fact, the voting results at hand can only be treated as a small sample. It is reasonable to assume that, for a review with helpfulness degree p ($0 \leq p \leq 1$) as the population proportion, each person would vote this review helpful with the probability of p independently. Therefore, the number of helpful votes X for a review with p and n can be regarded as a random variable following a binomial distribution, that is, $X \sim \text{bin}(n, p)$. According to Bayesian rule, the probability of “ x of n people found the review helpful” with p can be formulated as

$$f_{x|p}(x|p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \dots, n. \tag{1}$$

Then, the posterior distribution of p when “ x of n people found the review helpful” is observed can be calculated as

$$f_{p|x}(p|x) = \frac{f_{x|p}(x|p)f_p(p)}{\int_0^1 f_{x|p}(x|p)f_p(p)dp}, \tag{2}$$

where $f_p(p)$ is the prior distribution of p and often configured as a beta distribution [17].

The helpfulness distribution information on all observed reviews of concern can be incorporated by adjusting the parameters for the prior beta distribution of p . When no helpfulness distribution information can be observed or available, it is quite common to treat prior distribution of p as uniform distribution on $[0, 1]$ or beta distribution with parameters $(1, 1)$ [17]. Otherwise, the parameters (a, b) for prior beta distribution can be calculated as

$$a = \bar{p} \left(\frac{\bar{p}(1-\bar{p})}{\text{Var}(p)} - 1 \right), \quad b = (1-\bar{p}) \left(\frac{\bar{p}(1-\bar{p})}{\text{Var}(p)} - 1 \right), \tag{3}$$

where \bar{p} and $\text{Var}(p)$ are the sample mean and sample variance of observed reviews' percentage of helpful votes.

Thus, the posterior distribution of p is a beta distribution with parameters $(x + a, n - x + b)$, that is,

$$f_{p|x}(p|x) = \frac{[(x + a) + (n - x + b) - 1]!}{[(x + a) - 1]![(n - x + b) - 1]!} p^{(x+a)-1} (1 - p)^{(n-x+b)-1}. \quad (4)$$

Based on the posterior distribution of p , the expected review helpfulness value can be calculated as

$$\int_0^1 p f_{p|x}(p|x) dp = \int_0^1 p \frac{f_{x|p}(x|p) f_p(p)}{\int_0^1 f_{x|p}(x|p) f_p(p) dp} dp = \frac{x + a}{n + a + b}. \quad (5)$$

In this way, both the helpfulness voting result for a particular review and the helpfulness distribution on reviews with similar characteristics are considered. Notably, the expected review helpfulness degree may be viewed as a weighted sum of the traditional helpfulness estimation and the mean of prior beta distribution:

$$\frac{x + a}{n + a + b} = \frac{n}{n + a + b} \times \frac{x}{n} + \frac{a + b}{n + a + b} \times \frac{a}{a + b}. \quad (6)$$

In consideration of these two pieces of information, the proposed method can generate estimation results closer to real helpfulness, which will be shown in Sects. 3 and 4. Moreover, the ranking order of reviews for a particular product can be adjusted accordingly.

Furthermore, to cope with the limitation for confidence, the CI of estimated p could also be incorporated. Since reviews often receive helpful votes with extreme proportion (i.e., quite near 1 or 0) while the total number of votes is small, a way proven with more accuracy and narrower CI width is adopted [17] to find an minimum length interval $[c, d]$ (at given significance level α), satisfying the following criterion,

$$\int_c^d f_{p|x}(p|x) dp = 1 - \alpha. \quad (7)$$

Thus, the users' confidence on the helpfulness vote data (e.g., 9/10 vs. 90/100) could be reflected by the intervals (width), where narrower (broader) intervals mean higher (lower) confidence.

3 An Illustrative Example

For an online platform that allows users to share their opinions and experiences about products/services, suppose the helpfulness of all reviews of concern is observed. Then, the parameters for fitted beta distribution can be calculated by (3), and the expected helpfulness and related CI for a particular review can be estimated by (5) and (7), respectively.

In this context merely for illustrative purposes, the helpfulness voting information of Amazon reviews was collected. During June 21, 2012 to August 20, 2012, 624,940 reviews of 2,647 products, which were listed in the web pages in Full Store Dictionary [18], were collected. After filtering with total vote numbers no less than 100, the helpfulness distribution of reviews can be fitted with beta distribution at parameters (0.6559, 0.2825), as illustrated in Fig. 2. It is worth noticing that filtering with different numbers of minimum total votes (such as 10, 30, 50, 100, 300) could generate similar results.

Based on the helpfulness distribution observed from the real world, which was quite similar as distribution reported in [4], 5 reviews with random helpfulness degrees derived from the fitted beta distribution were constructed. For each review, the total helpfulness vote number was generated randomly from 1 to 10, so as to compare the performance of our proposed method with the traditional one under usual cases where total vote number was quite small. Then, each vote was assigned as a helpful vote with a probability equal to the helpfulness degree of this review. Finally, the reviews with their helpfulness degrees, total vote numbers, helpful vote numbers, and estimation results by both proposed and traditional methods were obtained and listed in Tables 1 and 2, separately.

Tables 1 and 2 show that the proposed method was able to estimate helpfulness based on helpfulness voting results more accurately. It can be seen that the proposed method is superior to the traditional one in light of estimation error, in that all five (in bold font) cases were lower. In addition, the proposed method was also

Fig. 2 Real helpfulness distribution with fitted beta distribution

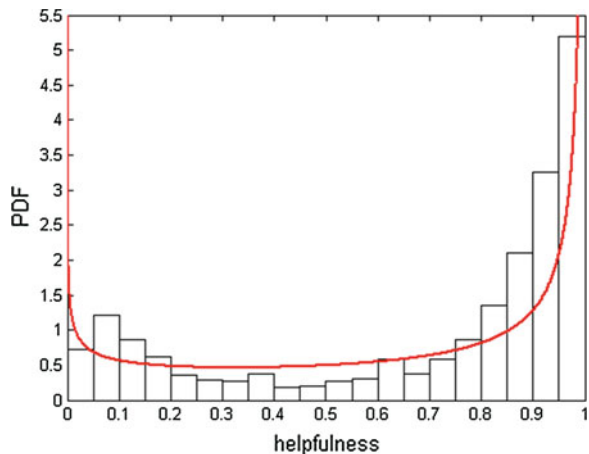


Table 1 Estimation results by traditional methods

Review ID	Real helpfulness	Observed results		Traditional estimation method	
		Helpful vote number	Total vote number	Estimated helpfulness	Estimation error
1	0.9698	8	8	1.0000	0.0302
2	0.7566	4	6	0.6667	0.0899
3	0.5436	2	5	0.4000	0.1436
4	0.3870	1	1	1.0000	0.6130
5	0.1237	0	9	0.0000	0.1237

Table 2 Estimation results by proposed methods

Review ID	Real helpfulness	Proposed estimation method			
		Estimated helpfulness	Estimation error	Estimated confidence intervals ($\alpha = 0.05$)	CI width ($\alpha = 0.05$)
1	0.9698	0.9684	0.0014	[0.7169, 1.0000]	0.2831
2	0.7566	0.6710	0.0856	[0.3154, 0.9195]	0.6041
3	0.5436	0.4472	0.0964	[0.1048, 0.7613]	0.6565
4	0.3870	0.1514	0.2356	[0.2236, 1.0000]	0.7764
5	0.1237	0.0660	0.0577	[0.0000, 0.2589]	0.2589

able to provide CIs which contain the real helpfulness degrees and can distinguish the reviews with the same helpfulness degrees (e.g., 9/10 vs. 90/100).

4 Experiments on Synthetic and Real Data

Experiments on both synthetic and real data were conducted to compare the estimation accuracy of the traditional and proposed methods, in the contexts of product type (e.g., books).

For synthetic data, the experimental environment was a PC platform with Intel i3-2100 CPU @ 3.10 GHz and 3.09 GHz, 2.94 GB Memory, Windows XP and MATLAB R2011b. For each synthetic experiment, there were 1,000 products of the same type with 1,000 reviews (each being synthesized with the total vote number n assigned as a random value following a uniform distribution on [10, 100]). The lower bound was set as 10 because many helpfulness prediction methods only adopted the reviews with no less than 10 votes as training data [6, 9, 16]; thus, the traditional estimation method was mostly used on reviews with at least 10 votes. The upper bound was set to be 100 as a quite high number of votes for a review in reality. For a review, the real helpfulness value was also assigned by random functions in MATLAB, that is, `unifrnd()`, `normrnd()`, `betarnd()`, and `gamrnd()` with different parameters, respectively. Different helpfulness distributions were chosen, as shown in Figs. 3 and 4, so as to consider various patterns of possible helpfulness distributions.

Fig. 3 Helpfulness distributions for random helpfulness degree generation (part 1)

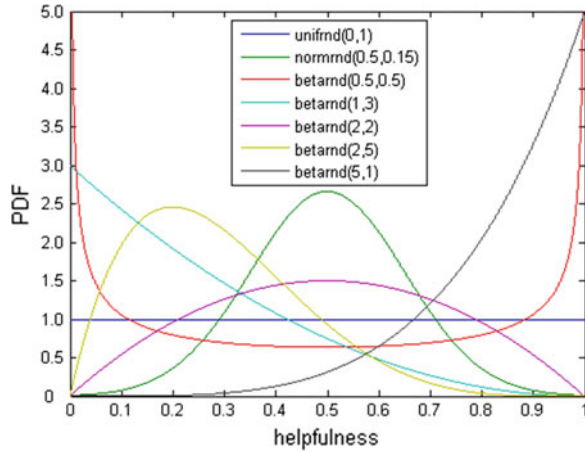
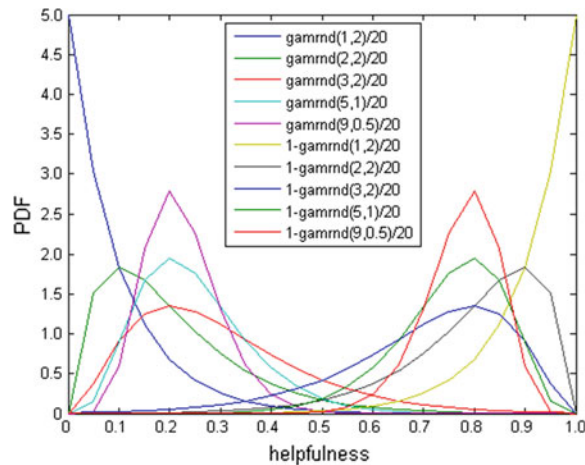


Fig. 4 Helpfulness distributions for random helpfulness degree generation (part 2)



After the real helpfulness and total vote number for a review were determined, each vote was assigned as a helpful vote with a probability equal to real helpfulness. Based on n votes, the helpful vote number for this review can also be observed as x , and both traditional and proposed methods were applied to estimating the helpfulness for this review.

To compare the estimation accuracy, the most popular measures in previous work [6, 9–11, 14, 15], that is, absolute error of estimated helpfulness and root mean squared error (RMSE), were used. Both of them are the smaller the better. The experimental results in Fig. 5 show that the proposed method outperformed the traditional method on both measures at all different helpfulness distributions.

Moreover, we also conducted experiments on real data, obtained from *book.dangdang.com*, which is one of the largest online book sellers in China. The review helpfulness voting data were collected for all reviews of 2,175 books (best

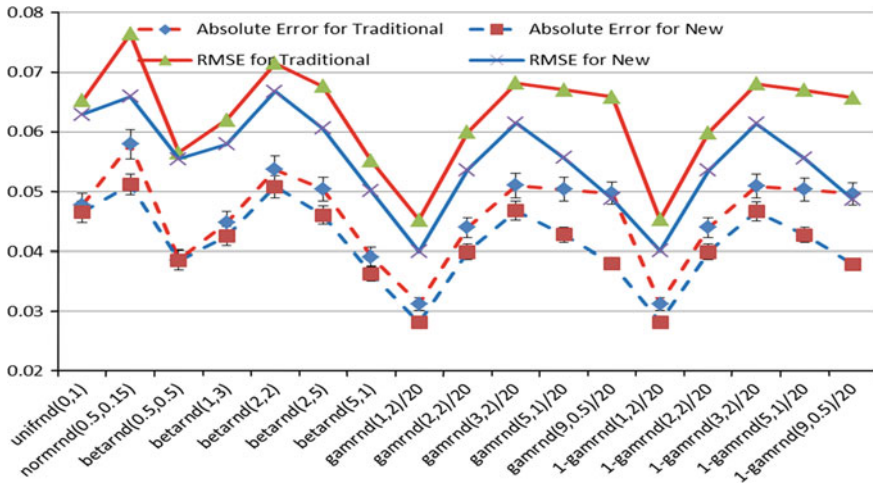


Fig. 5 Estimation accuracy of traditional and proposed methods on synthetic data

Table 3 Estimation accuracy of traditional and proposed methods on real data

Measures	Traditional method	Proposed method
Absolute error	0.01433 (var. = 0.0017)	0.01426 (var. = 0.0014)
RMSE	0.04403	0.03984

seller list from June 24, 2011 to September 6, 2011) in two time periods, that is, November 28, 2011—December 06, 2011 and December 29, 2011—January 09, 2012, amounting to a total of 1,699,313 reviews.

Since the helpfulness of a review can be defined as the proportion of population who vote this review helpful, the percentage of helpful votes at the 2nd time period for a review was regarded as the real helpfulness of this review when the total vote number was large enough, that is, this review received at least 100 votes at the 2nd time period and 10 more votes compared with the 1st time period. Furthermore, only those reviews with at least 10 votes at 1st time period were utilized to compare the estimation results since many helpfulness prediction methods only chose this number of reviews as their training data [6, 9, 16].

Finally, $a = 2.8453$ and $b = 0.3691$ were obtained by (3) based on helpfulness voting results at 1st time period. Estimation results by both traditional and proposed methods were listed in Table 3. It can be seen that the proposed method outperformed the traditional method on both measures.

5 Conclusion

Online review helpfulness estimation is a core step in online review helpfulness prediction. In this paper, a new method has been proposed to take into consideration the helpfulness distribution information of reviews, along with an ability to show the confidence on estimated helpfulness based on the estimated posterior distributions. An example has been provided to illustrate the effectiveness of the proposed method. Moreover, experiments on both synthetic and real data have been conducted, revealing the outperformance of the proposed method over the traditional one on estimation accuracy.

Future work can be carried out in developing new ways for both review helpfulness prediction and review ranking mechanism, as well as in applications with large online data for e-business services. Moreover, further investigations can also be made for online communities (e.g., consumer groups with different behaviors and tastes [19–21]), which may result in different helpfulness distributions.

Acknowledgments The work was partly supported by the National Natural Science Foundation of China (70890083/71072015/71110107027), Tsinghua University's Initiative Scientific Research Program (20101081741), and Research Center for Contemporary Management.

References

1. Pan Y, Zhang JQ (2011) Born unequal: a study of the helpfulness of user-generated product reviews. *J Retail* 87(4):598–612
2. Mudambi SM, Schuff D (2010) What makes a helpful online review? A study of customer reviews on amazon.com. *MIS Q* 34(1):185–200
3. Cao Q, Duan W, Gan Q (2011) Exploring determinants of voting for the “helpfulness” of online user reviews: a text mining approach. *Decis Support Syst* 50(2):511–521
4. Liu J, Cao Y, Lin C-Y, Huang Y, Zhou M (2007) Low-quality product review detection in opinion summarization. In: 2007 joint conference on empirical methods in natural language processing and computational natural language learning, 334–342
5. Siersdorfer S, Chelaru S, Nejdil W, Pedro JS (2010) How useful are your comments? Analyzing and predicting YouTube comments and comment ratings. In: 19th international conference on world wide web, ACM Press, North Carolina, 891–900
6. Liu Y, Huang X, An A, Yu X (2008) Modeling and predicting the helpfulness of online reviews. In: Eighth IEEE international conference on data mining, IEEE Press, 443–452
7. Kim S-M, Pantel P, Chklovski T, Pennacchiotti M (2006) Automatically assessing review helpfulness. In: 2006 conference on empirical methods in natural language processing, Association for computational linguistics, Sydney, 423–430
8. Ghose A, Ipeirotis PG (2011) Estimating the helpfulness and economic impact of product reviews: mining text and reviewer characteristics. *IEEE Trans Knowl Data Eng* 23(10):1498–1512
9. Yu X, Liu Y, Huang X, An A (2010) Mining online reviews for predicting sales performance: a case study in the movie domain. *IEEE Trans Knowl Data Eng* 24:720–734

10. Yu X, Liu Y, Huang X, An A (2010) A quality-aware model for sales prediction using reviews. In: 19th international conference on world wide web, ACM Press, North Carolina, 1217–1218
11. Zhang Z, Varadarajan B (2006) Utility scoring of product reviews. In: 15th ACM international conference on information and knowledge management, ACM Press, Virginia, 51–57
12. Tsur O, Rappoport A (2009) REVRANK: a fully unsupervised algorithm for selecting the most helpful book reviews. In: Third international ICWSM conference, 154–161
13. Wu PF, Heijden HVD, Korfiatis N (2011) The influences of negativity and review quality on the helpfulness of online reviews. In: International conference on information systems
14. Lu Y, Tsaparas P, Ntoulas A, Polanyi L (2010) Exploiting social context for review quality prediction. In: 19th international conference on world wide web, ACM Press, North Carolina, 691–700
15. Moghaddam S, Jamali M, Ester M (2011) Review recommendation: personalized prediction of the quality of online reviews. In: 20th ACM international conference on information and knowledge management, ACM Press, Scotland, 2249–2252
16. Danescu-Niculescu-Mizil C, Kossinets G, Kleinberg J, Lee L (2009) How opinions are received by online communities: a case study on amazon.com helpfulness votes. In: 18th international conference on world wide web, ACM Press, Madrid, 141–150
17. Timothy DR (2003) Accurate confidence intervals for binomial proportion and Poisson rate estimation. *Comput Biol Med* 33(6):509–531
18. Full Store Directory (2012) http://www.amazon.com/gp/site-directory/ref=sa_menu_fullstore
19. Hennig-Thurau T, Gwinner KP, Walsh G, Gremler DD (2004) Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet? *J Interact Marketing* 18(1):38–52
20. Wu F, Huberman BA (2010) Opinion formation under costly expression. *ACM Trans Intell Syst Technol* 1(1):1–13
21. Otterbacher J (2011) Being heard in review communities: communication tactics and review prominence. *J Comput Mediated Commun* 16(3):424–444

Probabilistic Attribute Mapping for Cold-Start Recommendation

Guangxin Wang and Yinglin Wang

Abstract Collaborative filtering recommender system performs well when there are enough historical data of the users' online behavior, but it does not work on new users who have not rated any items, or new items that have not been rated by any users, which are called cold-start user and cold-start item, respectively. In order to alleviate the cold-start problem, additional information such as the attributes of users and items must be used. We propose a novel hybrid recommender system, which tries to construct the probabilistic relationship between user attributes and movie attributes using EM algorithm. It can make recommendation for both new users and new items. We evaluate our approach on MovieLens dataset and compare our method with the state-of-the-art approach. Experimental results show that the two approaches have almost the same performance, while our approach uses less time to train the model and make online recommendation.

Keywords Recommender system · Cold start · User and item features · Latent variable model

1 Introduction

Recommendation systems aim at recommending personalized information to users based on the historical data, so as to improve user experience. They are playing an important role in today's online markets. Examples include Amazon.com [1], which recommends CD, books, and other products to customs, and MovieLens [2],

G. Wang (✉) · Y. Wang
School of Computer Science and Engineering, Shanghai JiaoTong University,
Shanghai, China
e-mail: wxin@sjtu.edu.cn

Y. Wang
e-mail: ylwang@sjtu.edu.cn

which recommends movies. Several methods have been proposed to make better recommendation. Generally, these methods can be divided into two categories, content-based filtering and collaborative filtering.

Content-based filtering recommends an item to a user based on a description of the item and the profile of the user's interest. The profile of a user is created and updated automatically when an item is rated by the user [3]. One of the advantages of content-based recommendation is that it can recommend users' new items. However, it cannot recommend to new users because the profiles of new users have not been created. What is more, content-based filtering suffers from "over-specialization" problem [4]. That is, the system may recommend to a user items that are similar to those already rated by the user. For example, in a movie recommendation system, if the user only rated comedy movies, then the system may not recommend other genre movies, such as love stories, to the user. Finally, content-based filtering is limited by the feature selection problems. They have to use different feature selection algorithms with different domains.

Unlike content-based filtering, collaborative filtering provides recommendation based on the past user behavior, including explicit and implicit feedbacks, such as rating data and purchase history. Collaborative filtering overcomes some disadvantages of content-based filtering. First, it can recommend items with very different content, as long as other users show their interest in these items. Second, collaborative filtering is domain free. That is, it treats each item with the associated ratings made by users, regardless of what the item is. So it has no feature selection problems.

Though collaborative filtering performs well when there are enough historical data, it suffers from serious cold-start problem. It would not be able to recommend new user or new item until the new user has rated enough items, or the new item has been rated by enough users. In order to solve the cold-start problem, we proposed a new hybrid approach that utilizes not only user ratings but also the attributes associated with users and items. We exploit the probabilistic latent semantic analysis (pLSA) [5] to model the relationship between user attributes and item attributes, and use expectation maximization (EM) [6] algorithm to fit the model. We evaluated our approach on the MovieLens dataset and compared our approach with the state-of-the-art approach. We also evaluated time efficiency of our approach. Experimental results show that our approach can make more accurate recommendation while using less time.

This paper is organized as follows. [Section 2](#) describes our algorithm. [Section 3](#) describes our experiments. [Section 4](#) surveys related works on cold-start recommendation. In [Sect. 5](#), we give our conclusion and future work.

2 Methods

In this section, we will describe our probabilistic attribute-to-attribute mapping (PATAM) approach to solve the cold-start collaborative filtering, by exploiting the probabilistic latent semantic analysis model.

2.1 The Probabilistic Latent Semantic Analysis Model

The probabilistic latent semantic analysis model (pLSA for short) was proposed in [5, 7]. It was first used to model the co-occurrence of documents and words. Then in [5], it is used to model the co-occurrence of user–item pairs in collaborative filtering.

In the model, it is assumed that there exists a latent variable z for each user–item pair (u, y) , which motivates user u to prefer item y because of z . It is assumed that user u and item y are conditionally independent given z . The pLSA model is a generative model, meaning that each user–item pair (u, y) is generated by the following procedure.

1. Select a user with the prior probability of $P(u)$.
2. Pick a latent variable z with conditional probability of $P(z|u)$.
3. Generate a item y with conditional probability of $P(y|z)$.

The joint distribution over (u, y) is:

$$P(u, y) = \sum_z P(u) * P(z|u) * P(y|z) \tag{1}$$

To maximize the log-likelihood, parameters are calculated using EM algorithm.

2.2 Probabilistic Attribute-to-Attribute Mapping

The recommendation system described above is a pure collaborative filtering model, so it cannot deal with the cold-start problem. Inspired by content-based filtering, we utilize the user and item attributes in our model to help us make better recommendation when there are no rating data available.

In this case, each user is represented by a set of attributes, denoted as a boolean vector α , $\alpha \in R^N$, where N is the number of user attributes. Similarly, each item is represented by a set of attributes, denoted as a boolean vector β , $\beta \in R^M$, where M is the number of item attributes. Both vector α and vector β are boolean vectors, that is, if a user has the i th attribute, then $\alpha_i = 1$, else $\alpha_i = 0$. Usually, user attributes consist of the age, gender, occupation, location, and other useful information. The attributes of items depend on the type of items. For example, in a movie recommendation system, the possible attributes are genre, actors, and director of movies, but in a commodity recommendation system, the useful attributes may be the application of the commodity and so on.

In order to model the relationship between the N user attributes and M item attributes, we use the pLSA model described above.

Given a user–item rating dataset, from the viewpoint of pLSA model, the whole dataset is generated as

$$D = \prod_{i=1}^N \prod_{j=1}^M P(\alpha_i, \beta_j)^{n(\alpha_i, \beta_j)}. \tag{2}$$

In Eq. (2), $n(\alpha_i, \beta_j)$ denotes the number of co-occurrence of α_i and β_j . In order to utilize the explicit user ratings, we define $n(\alpha_i, \beta_j)$ as

$$n(\alpha_i, \beta_j) = \sum_{u \in U(\alpha_i)} \sum_{y \in Y(\beta_j)} R(u, y) \tag{3}$$

where $U(\alpha_i)$ denotes the set of users who has attribute α_i , and $Y(\beta_j)$ denotes the set of items which has attribute β_j , and $R(u, y)$ denotes the rating user u rated to item y .

The log-likelihood L of the whole dataset is

$$L = \log D = \sum_{i=1}^N \sum_{j=1}^M n(\alpha_i, \beta_j) \log(P(\alpha_i, \beta_j)) \tag{4}$$

To maximize the log-likelihood L , we use the EM algorithm as in [5].

The steps of EM algorithm are as follows:

E-Step:

$$P(z|\alpha_i, \beta_j) = \frac{P(z)P(\alpha_i|z)P(\beta_j|z)}{\sum_{z'} P(z')P(\alpha_i|z')P(\beta_j|z')}$$

M-Step:

$$\begin{aligned} P(\beta_j|z) &= \frac{\sum_{\alpha_i} n(\alpha_i, \beta_j) P(z|\alpha_i, \beta_j)}{\sum_{\alpha_i, \beta'_j} n(\alpha_i, \beta'_j) P(z|\alpha_i, \beta'_j)} \\ P(\alpha_i|z) &= \frac{\sum_{\beta_j} n(\alpha_i, \beta_j) P(z|\alpha_i, \beta_j)}{\sum_{\alpha'_i, \beta_j} n(\alpha'_i, \beta_j) P(z|\alpha'_i, \beta_j)} \\ P(z) &= \frac{1}{R} \sum_{\alpha_i, \beta_j} n(\alpha_i, \beta_j) P(z|\alpha_i, \beta_j) \\ R &= \sum_{\alpha_i, \beta_j} n(\alpha_i, \beta_j) \end{aligned}$$

To make recommendation for a user, we calculate the conditional probability $P(\beta|\alpha)$ between this user and all the items. Since we assume that item’s attributes are independent and identically distributed (i.i.d), and user’s attributes are also i.i.d,

$$\begin{aligned} P(\beta|\alpha) &= P(\beta_1, \beta_2, \dots, \beta_M | \alpha_1, \alpha_2, \dots, \alpha_N) \\ &= \prod_{j=1}^M P(\beta_j | \alpha_1, \alpha_2, \dots, \alpha_N) \\ &\propto \prod_{j=1}^M P(\beta_j) \prod_{i=1}^N P(\alpha_i | \beta_j) \\ &= \prod_{j=1}^M \frac{\prod_{i=1}^N P(\alpha_i, \beta_j)}{P(\beta_j)^{N-1}} \end{aligned}$$

Similarly, to make recommendation for a new item, we can calculate $P(\alpha|\beta)$ as

$$\begin{aligned}
 P(\alpha|\beta) &= P(\alpha_1, \alpha_2, \dots, \alpha_N | \beta_1, \beta_2, \dots, \beta_M) \\
 &= \prod_{i=1}^N P(\alpha_i | \beta_1, \beta_2, \dots, \beta_M) \\
 &\propto \prod_{i=1}^N P(\alpha_i) \prod_{j=1}^M P(\beta_j | \alpha_i) \\
 &= \prod_{i=1}^N \frac{\prod_{j=1}^M P(\alpha_i, \beta_j)}{P(\alpha_i)^{M-1}}
 \end{aligned}$$

In the above equations, there are too many probability multiplications, so it is easy to get a zero result. So in reality, we use $\log P(\beta|\alpha)$ and $\log P(\alpha|\beta)$ to rank items and users.

3 Experiments

In this section, we will evaluate our proposed approach on the MovieLens 1 M dataset, which is a widely used dataset. We did not use Netflix dataset because it does not contain any user attributes.

3.1 Compared Method

We compared our approach with the pairwise preference regression method described in [8] and the user–actor aspect model described in [9].

3.1.1 The Pairwise Preference Regression Method

Just as our approach, this method can also make recommendation for new users and new items. It is a regression model that optimizes the following loss function

$$\arg \min_w \sum_{ui \in O} (r_{ui} - s_{ui})^2 + \lambda \|w\|_2^2 \tag{5}$$

In Eq. (5), r_{ui} is the real rating between user u and item I , and s_{ui} is the predict rating defined by $s_{ui} = \sum_{a=1}^C \sum_{b=1}^D x_{u,a} z_{i,b} w_{a,b}$. \vec{x} and \vec{z} are user attribute vector and item attribute vector, respectively; C and D are the dimensionality of \vec{x} and \vec{z} respectively.

3.1.2 The User–Actor Aspect Model

The user–actor aspect model also tries to model the relationship between users and the actors of movies. It tries to fit the joint distribution of users and actors using the following probability.

$$P(u, a) = \sum_z P(z)P(u|z)P(a|z) \quad (6)$$

It also uses the EM algorithm to train the model.

E-Step:

$$P(z|a, i) \propto P(a|z)P(z|i)$$

M-Step:

$$P(z|i) \propto \sum_a n(a, i)P(z|a, i)$$

In order to recommend new items to users, items are ranked according to $P(u|i)$

$$P(u|i) = \sum_z P(u|z)P(z|i)$$

Since it only takes the attribute of items into consideration, it cannot make recommendation for new users.

3.2 Evaluation Metrics

We use mean average precision (MAE) to measure the performance of our approach, which is a widely used evaluation measure in information retrieve and has been shown to have especially good discrimination and stability [10]. For each query, it first calculates the precision at different recall levels, then calculates the average precision (AP) as follows:

$$\text{AveP} = \frac{\sum_{k=1}^n [P(k) \times \text{rel}(k)]}{\text{number of relevant documents}} \quad (7)$$

where $P(k)$ is the precision at recall level k , $\text{rel}(k)$ is an indication function, and $\text{rel}(k) = 1$ when the item at rank k is a relevant document, $\text{rel}(k) = 0$ otherwise.

MAP for a set of queries is the mean of average precision score of each query.

$$\text{MAP} = \frac{\sum_{q=1}^Q \text{AveP}(q)}{Q} \quad (8)$$

where Q is the number of queries.

When evaluating recommendation system, we treat a user as a query. Movies that are rated by a user and rating value ≥ 4 are treated as relevant documents. We calculate the average precision for each test user and then calculate the MAP for all the test users.

3.3 Testing Methodology

We use the benchmark dataset MovieLens. The MovieLens 1 M dataset contains 1 million anonymous ratings of 3,706 movies by 6,040 users. Each rating is between 1 and 5 inclusive.

In MovieLens 1 M dataset, each user has three attributes: gender, age, and occupation. Each movie has two attributes: movie genre and release time. Similar to [8], we categorized the movie release year into 8 groups, that is, $\geq 2,000$, 90s, 80s, 70s, 60s, 50s, 40s, and <1940 , and the age of users is categorized into 7 groups, that is, under 18, 18–24, 25–34, 35–44, 45–49, 50–55, and above 56. In order to make better recommendation, we add additional information from the internet movie database (IMDB). The attributes we used in our experiment are summarized in Table 1.

We first remove those movies which have no director information and corresponding ratings. Similar to [8], we evaluate our approach in three scenarios, that is, cold-start users, cold-start movies, and cold-start users and movies. We randomly divided the users into 1,005 test users and 5,035 training users. Similarly, we randomly divided the movies into 823 test movies and 1,508 training movies. As a result, we divided the dataset into four parts: one part for training the model and the other three parts for testing the three scenarios.

Since the user–movie rating matrix of the training dataset is very sparse (the training dataset contains 1,508 movies and 5,035 users, but only 6,28,738 ratings, and the average density ratio is 0.0828), it is hard to reveal the relationship between user attributes and movie attributes. What is more, it may lead to serious overfitting. In order to alleviate the problem, we propose to use a collaborative filtering algorithm to predict the missing ratings before making cold-start recommendation. In fact, any collaborative filtering algorithm can be applied here. In our experiment, we use the item-based collaborative filtering recommendation algorithm described in [11].

Table 1 The user attributes

	Attribute name	Number
User Attributes	Gender	2
	Age	7
	Occupation	21
Movie Attributes	Genre	18
	Release time	8
	Director	1,360

3.4 Experimental Result

Figure 1 shows the MAE score comparing our approach with the pairwise regression model in three scenarios

Our approach outperforms the pairwise regression model both in cold-start user scenario and in cold-start user and movie scenario. However in another scenario, pairwise regression model gives a little more accurate prediction than our model, but the gap is very small. As reported in [8], cold-start user scenarios have better performance than cold-start movie scenario. This may be because the selected attributes of users are more representative and discriminative. The user-actor aspect model can only work in cold-start movie scenario, and the MAE is lower than PATAM and pairwise regression.

The left part of Fig. 2 shows the time used to train our model, pairwise regression model, and the user-actor aspect model on the same training dataset. The right part of Fig. 2 shows the average time used to make recommendation for a user. We can see that our model needs much less time than pairwise regression

Fig. 1 MAE score on three scenarios

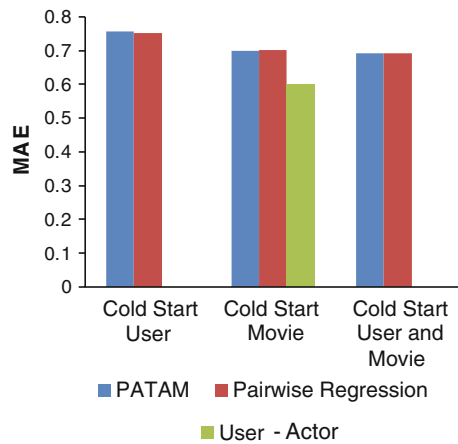
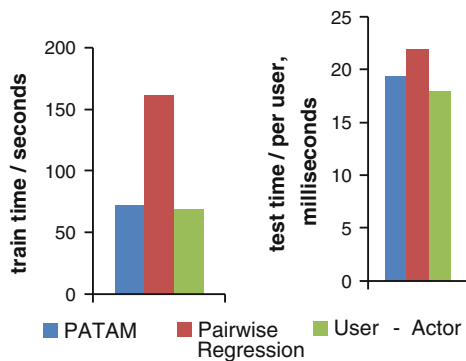


Fig. 2 The left part shows the time used to train each model. The right part shows the time used to make recommendation to a user



model in both situations since the latter needs many matrix manipulations. The user–actor aspect model uses less time than our model, but the gap is not obvious.

4 Related Work

There are mainly two approaches to build recommender system, that is content-based filtering and collaborative filtering. Content-based filtering recommends items similar to those items the users have explicitly or implicitly shown their interest in. Collaborative filtering uses community information. Some representative collaborative filtering approaches are k nearest neighbors [11–13] and matrix factorization [14–16].

Content-based filtering suffers from lack of diversity, whereas collaborative filtering can overcome this problem, but it suffers from the cold-start problem.

Several studies have been done to deal with the cold-start problem, such as [17–20] by constructing an interview process, by choosing several items to present to new users, and by collecting their ratings, in order to learn a new user’s preference. The critical problem is how to choose items for users to rate, so as to minimize a new user’s rating effort and, at the same time, get more useful information to make smart recommendation. Rashid et al. [18] investigated several statistics criteria, such as popularity, contention, pure entropy. Golbandi et al. [20] argued that it is unreasonable to assemble multiple criteria. They proposed a method called “GreedyExtend”, which explicitly optimizes prediction accuracy during the construction of seed set [20]. These methods generate item set statically in advance, and the item set will not change during the whole interview process. That is, each new user is presented the same item set to rating, regardless of what the user’s feedback is during the interview process. As Rashid et al. [18] pointed out, these methods lack personalization. So, later works, such as [17, 19], use decision tree to personalize the interview process.

Another effective approach to overcome the cold-start problem is to exploit users’ and items’ available attributes. Popescul et al. [21] extended Hofmann’s two-way aspect model [5] to a three-way aspect model, by modeling the co-occurrence data among users, items, and item content. Schein et al.’s [9] model is also based on Hofmann’s two-way aspect model. They proposed person/actor model, instead of user/item model, to combine user and item content under a single probabilistic framework. However, it only takes one attribute of movies into consideration, which is too limited to reveal the real interest of users. What is more, it tries to model the relationship between users and actors, instead of the attributes of users. So it can only make recommendation for new items. These methods only take item’s attributes into account, so they do not work for new user recommendation. Gantner et al. [22] map users’ and items’ attribute to latent feature space, so as to construct an attribute-aware matrix factorization model. Park et al. [8] used a regression model that optimizes pairwise preferences. These methods can recommend new items and new users.

5 Conclusion and Future work

We have proposed PATAM, a hybrid approach that exploits the attributes of users and items for cold-start recommendation. We use the pLSA model to model the joint distribution between user attributes and item attributes, and use EM algorithm to train the parameters in the model. Experiments on MovieLens dataset show that our approach outperforms the state-of-the-art methods and, at the same time, uses less time to train the model and make recommendation.

In the future, we will try to take contextual information into account, which is called context-aware recommendation. Potential contextual information includes the current location of users, and the time recommendation is made to a user and so on.

Acknowledgments This paper was supported by the Science and Technology Innovation Action Plan (Grant Number:12511502902) of Shanghai Science and Technology Committee.

References

1. Linden G, Smith B, York J (2003) Amazon.com recommendations: item-to-item collaborative filtering. *Internet Comput IEEE* 7(1):76–80
2. Miller BN, et al (2003) MovieLens unplugged: experiences with an occasionally connected recommender system. In: *Proceedings of 8th international conference on intelligent user interface*, ACM 2003
3. Pazzani M, Billsus D (2007) Content-based recommendation systems. In: Brusilovsky P, Kobsa A, Nejdl W (eds) *The adaptive web*. Springer, Berlin, pp 325–341
4. Desrosiers C, Karypis G (2011) A comprehensive survey of neighborhood-based recommendation methods. In: Ricci F (ed) *Recommender systems handbook*, Springer, Berlin, pp 107–144
5. Hofmann T (1999) Probabilistic latent semantic indexing. In: *Proceedings of 22nd annual international ACM SIGIR conference on research and development in information retrieval*, ACM
6. Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Ser B (Method)*, 1977:1–38
7. Hofmann T (2001) Unsupervised learning by probabilistic latent semantic analysis. *Machine Learn* 42(1):177–196
8. Park ST, Chu W (2009) Pairwise preference regression for cold-start recommendation. In: *Proceedings of the 3rd ACM conference on recommender systems*, ACM
9. Schein AI, et al (2002) Methods and metrics for cold-start recommendations. In: *Proceedings of the 25th annual international ACM SIGIR conference on research and development in information retrieval*, ACM
10. Manning CD, Raghavan P, Schütze H (2008) *Introduction to information retrieval*, Vol 1. Cambridge University Press, Cambridge
11. Sarwar B, et al (2001) Item-based collaborative filtering recommendation algorithms. In: *Proceedings of the 10th international conference on world wide web*, ACM
12. Shardanand U, Maes P (1995) Social information filtering: algorithms for automating “word of mouth”. In: *Proceedings of the SIGCHI conference on human factors in computing systems*

13. Herlocker JL, Konstan JA, Borchers A, Riedl J (1999) An algorithmic framework for performing collaborative filtering. In: Proceedings of the 22nd annual international ACM SIGIR conference on research and development in information retrieval
14. Koren Y, et al (2009) Matrix factorization techniques for recommender systems. *Computer* 42:30
15. Jamali M, Ester M (2010) A matrix factorization technique with trust propagation for recommendation in social networks. In: Proceedings of the 4th ACM conference on recommender systems RecSys'10, ACM
16. Yehuda Koren. Collaborative filtering with temporal dynamics. In: Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining, Communications of the ACM
17. Zhou K, Yang SH, Zha H (2011) Functional matrix factorizations for cold-start recommendation. In: Proceedings of the 34th international ACM SIGIR conference on research and development in information retrieval, ACM
18. Rashid AM, et al (2002) Getting to know you: learning new user preferences in recommender systems. In: Proceedings of the 7th international conference on intelligent user interfaces, ACM
19. Golbandi N, Koren Y, Lempel R (2011) Adaptive bootstrapping of recommender systems using decision trees. In: Proceedings of the 4th ACM international conference on web search and data mining, ACM
20. Golbandi N, Koren Y, Lempel R (2010) On bootstrapping recommender systems. In: Proceedings of the 19th ACM international conference on information and knowledge management, ACM
21. Popescul A, et al (2001) Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments. In: Proceedings of the 17th conference on uncertainty in artificial intelligence
22. Gantner Z, et al (2010) Learning attribute-to-feature mappings for cold-start recommendations. In: Data mining (ICDM) 2010 IEEE 10th international conference

Measuring Landscapes Quality Using Fuzzy Logic and GIS

Victor Estévez González, Luis Garmendia Salvador
and Victoria López López

Abstract A methodology to evaluate visual quality and fragility of landscapes with fuzzy logic and GIS is given. The fuzzy concept of landscape is modeled with fuzzy sets and fuzzy connectives and used in ArcGIS to evaluate every point of a map.

Keywords GIS · Fuzzy logic · Visual fragility · Visual quality · Landscape · Intrinsic visual fragility · Acquired visual fragility

1 Introduction

Landscape is one of the natural resources that nowadays have a greater ecological and social demand. Proper landscape management requires justified actions and to set adapted actuations to the environment without changing or degrade his character.

A methodology for analyzing and evaluating the landscape allows to draw useful conclusions for visually integrating the activities in its territorial context, whether inclusion in a management plan or simply when implementing an industry in a place where their visual impact is minimal, in addition to being one of the most important factors when making an assessment of environmental impact.

V. E. González (✉)

MTIG Geography, Complutense University of Madrid, Madrid, Spain

e-mail: vstevez@gmail.com

L. G. Salvador

DISIA, Computer Science Faculty, Complutense University of Madrid, Madrid, Spain

e-mail: lgarmend@fdi.ucm.es

V. L. López

DACYA, Computer Science Faculty, Complutense University of Madrid, Madrid, Spain

e-mail: vlopez@fdi.ucm.es

To evaluate the landscape in this work, the visual quality and fragility are objectively assessed. The quality of the landscape depends on the uses and activities that develop. The quality is more subjective than visual fragility, which takes into account the amount of area viewed by potential observers.

The exponential development of GIS software has been a mainstay in many fields as geomarketing, accessibility plans, environmental management, network management, transportation and cadastre and land management.

This type of analysis would be unthinkable without the power of GIS analysis.

1.1 Objectives

This paper aims to develop a methodology to integrate the assessment of landscape processes such as physical planning and land use planning and trying to avoid the subjectivity of landscape concept. This subjectivity is present in most related papers. This one incorporates weighted viewshed to gain objectivity in the analysis.

To achieve this, we use multicriteria evaluation with fuzzy logic integrated into the ArcGIS software that allows us to apply this methodology to different study areas.

1.2 Study Areas

Two different areas have been chosen to apply this methodology:

Sanxenxo (Galicia) was chosen because it is a coastal town, so it presents high values of landscape quality. Anthropogenic pressure generated in this county is minimal. It has a population of 17,586 people, which in its 44 km² implies a population density of 400 km².

San Fernando de Henares (Madrid) was chosen due to the proximity of major roads and towns. Unlike in the Galician town, the pressure on this county is much higher, almost doubling its population in the same surface.

2 Preliminaries

2.1 Visual Quality and Fragility

Developments by several authors [1–6] have been chosen to develop a model able to integrate the visual quality, visual intrinsic fragility, and visual acquired fragility of the landscape.

The perception of the visual quality of the landscape is a creative act of interpretation made by the observer [7]. The territory has intrinsic qualities in their natural or artificial elements that are perceived by each of the observers of the territory. This means that the visual quality of the landscape is appreciated and recognized differently depending on the profile of each observer.

For the evaluation of the visual quality, we have considered the proposal of [8] that considers the visual quality formed by

- Diversity: It quantifies the degree of mosaic of uses in the landscape. Greater landscape diversity has higher quality.
- Ecological value: The proximity to areas of great ecological value (forest of oaks, riparian forests, etc.) has higher quality.
- Naturalness: The more natural landscape is more susceptible to damage. The most natural landscape has the highest value.
- Proximity to high value zones: Proximity to these zones increases the quality of the adjacent landscape.
- Proximity to low value zones: Proximity to a visual impact reduces the quality of the adjacent landscape.

The visual fragility is the answer to use of it, the degree of damage able to change their properties. The opposite is the visual absorption [9] understood as the ability to receive changes without reducing the visual quality. The greater fragility visual the lower absorption capacity.

To evaluate the fragility, we propose a method inspired by [9] and [10], which distinguishes between intrinsic and acquired fragility:

- Visual intrinsic fragility: A territory has its own characteristics and properties, for example slopes, orientations, land.
- Visual acquired fragility: A territory viewed by observers, both mobile (roads) and fixed (population centers). For the calculation takes into account, the proposal of [8] performing analysis using cumulative viewshed.

2.2 ArcGIS Fuzzy Module

You can perform the analysis by introducing fuzzy logic for each landscape parameter with the analyst module of ArcGIS 10.0 Spatial. There are two tools for the implementation of the fuzzy logic.

Fuzzy membership [12] allows assignment of a parameter to a fuzzy membership function that can be defined by the type of function. This tool has a set of functions to perform the assignment of membership degrees to predicates.

Fuzzy overlay is used to merge the fuzzy membership. The module allows the connection of multiple criteria by different operators of conjunction, disjunction, and fuzzy aggregation [11].

3 Sources and Methodology

3.1 Employed Sources

The geographic information necessary in this paper is property of the National Geographic Institute of Spain; it is free for non-commercial works.

All used data are in the ETRS 1989 system.

3.2 Methodology of Landscape Measuring

These are the parameter to measure the visual quality and its fuzzy function:

- Naturalness: Lineal from 1 to 5. High values have more possibility to belong to the fuzzy set.
- Distance to high naturalness zones: MS small with the midpoint based on its mean and standard deviation. Proximity to high naturalness zones has more possibility to belong to the fuzzy set.
- Distance to low naturalness zones: MS large with the midpoint based on its mean and standard deviation. Proximity to low naturalness zones has less possibility to belong to the fuzzy set.
- Diversity: Lineal from 1 to 5. High values have more possibility to belong to the fuzzy set.
- Chromatic contrast: Lineal from 1 to 5. High values have more possibility to belong to the fuzzy set.
- West orientation: Gaussian with midpoint 270. It keeps a normal distribution, the closer the value 270 more possibility to belong to the fuzzy set.

All parameters are reclassified from 0 to 1. With the OR disjunction operation of fuzzy overlay, we get the visual quality.

Once we have the visual quality, the visual fragility is computed as the fuzzy overlay between visual acquired fragility and visual intrinsic fragility.

Visual intrinsic fragility has these components adapted to fuzzy logic:

- Aspect: Gaussian with midpoint 270. It keeps a normal distribution, the closer the value 270 more possibility to belong to the fuzzy set.
- Slope: MS large with the midpoint based on its mean and standard deviation. High values have more possibility to belong to the fuzzy set. High values produce a greater visual impact.
- Land use: Lineal from 1 to 5. High values have more possibility to belong to the fuzzy set. High values produce a greater visual impact.

Visual acquired fragility is computed by the fuzzy overlay between acquired visual fragility from fixed and mobile observers.

The fixed points are the localities. We use the collected population centers in the latest Census of Population and Gazetteer with influence in the study area. The value of the visual acquired fragility of fixed observer is the sum of the viewshed of “n” fixed points that depends on the distance to the center (*d*), the specific viewshed (*CV*), and population registered of the fixed (*P*):

$$FVA_{obs.fijos} = \sum_{i=1}^n (d_i * CV_i * P_i) \tag{1}$$

The mobile points are points on the roads of the study area; therefore, the methodology is very similar, and instead of population, we have average daily traffic:

$$FVA_{vias} = \sum_{j=1}^n (d_j * CV_j * T_j)$$

The final zoning is computed by the fuzzy overlay between the visual quality and visual fragility. The result is a raster with values between zero and one. High values produce a greater visual impact.

Once the result is computed, the defuzzification is implemented, so we select values over a critical value, in this case 0, 7. This new reclassification indicates the most vulnerable areas.

4 Results

The tables show the expected results, more anthropic pressure generated in San Fernando than in Sanxenxo; however, this county presents higher values of visual quality.

4.1 Sanxenxo Map

The following chart represents the percentage of covert area of each category; the results are reclassified into five categories for better comparison (Table 1).

In this county of 4.393 ha, an area of 86.420 ha has been studied; 229.340 fixed observers and 492.700 mobile observers have been counted.

Table 1 Summary percentage hectares of each category for every variable in Sanxenxo, using fuzzy logic

	FVI (%)	FVA (%)	FV (%)	Calidad (%)	Zonificación (%)
1	27	83	27	0	11
2	27	9	19	58	21
3	18	2	20	28	13
4	14	1	18	11	25
5	13	5	16	3	29

4.2 San Fernando de Henares Map

In this county of 3.986 ha, an area of 87.218 ha has been studied. Studying the acquired visual fragility, 815.097 fixed observers and 492.700 mobile observers have been counted (Table 2).

5 Results

5.1 Fuzzy Logic

The results shown in the charts show that there is more anthropic pressure generated in San Fernando than in Sanxenxo; however, this land presents higher values of visual quality. The high values are due to the disjunction operations on the sets. Only one parameter with high values produces a high value in the final visual fragility or quality, so with high values in some parameters produce a very high value of visual fragility or quality.

Fuzzy logic can model and operate information with uncertainty or without a clear definition. In this case, assessing the fuzzy sets in each feature of the pixel universe in the study zone and evaluating from them, more developed concepts with logical connectives. Clear concept in these operations could also appear with

Table 2 Summary percentage hectares of each category for every variable in San Fernando De Henares using fuzzy logic

	FVI (%)	FVA (%)	FV (%)	Calidad (%)	Zonificación (%)
1	63	17	12	60	10
2	3	13	10	15	7
3	11	5	8	8	6
4	5	27	20	4	10
5	7	28	40	2	56

value zero or one. Fuzzy logic can process data as a human mind does. The principal advantage is the way it can adjust the reality of the territory (slope, aspect, elevations) to raster format, very useful in, for example, searching for potential planting areas or similar.

A defuzzification method is used to make a final decision. A binary map is obtained with fragile area or not fragile area without complex categories.

Future works can choose different t-conorms [11] operators for the disjunction OR, that is, the maximum or probabilistic sum instead of the SUM, and also several t-norms can also be chosen for the operator AND define predicates from other requisites. It would be possible to do a further fuzzy sets modeling of predicates.

5.2 On Methodology

This paper has shown that it is possible to consider complex parameters, as visual quality and fragility establish a zoning that takes into account landscape variables in process of landscape management or assessment of environmental impact.

The proposed model can, with minimal change, be applied to greater areas than in this paper, or particular cases of projects or concrete works.

The visual fragility computed by the fuzzy overlay of all the viewsheds reduces subjectivity, besides increasing the consistence, of the results without supposing an additional cost.

Consistent and comparable results obtained in two areas with differences in topography, morphology, land use, and human pressure indicate that the model could be applied in any area.

The key has been to minimize the subjectivity inherent in operator, present in any landscape assessment process. Thus, the procedure may be homologous, and their results should be consistent regardless of the operator to apply the methodology.

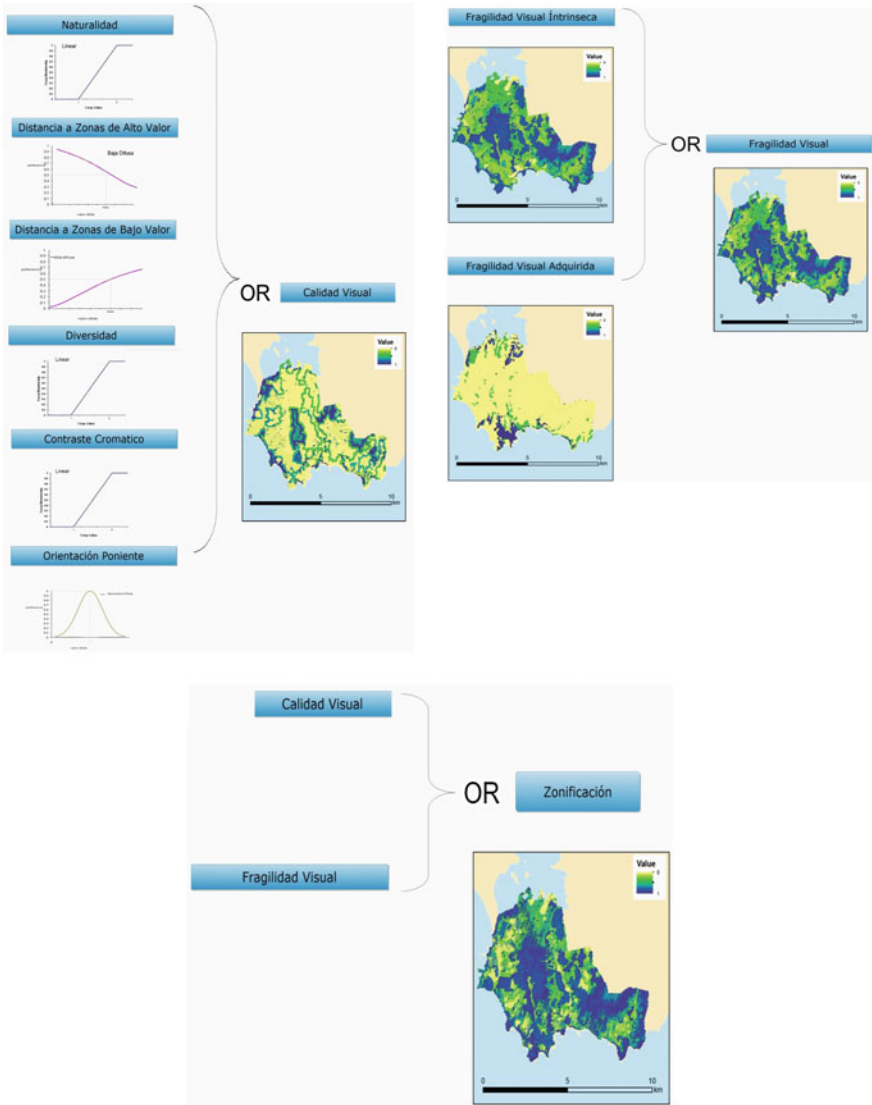
The model can easily evolve adopting new parameters as the phenomena of concealment, past trends in land use, etc.

The resulting model is sufficiently transparent to promote public participation processes, and the results obtained have no discretion and may audit, verify, and justify what are the factors that affect the classification of a given area from a landscape point of view.

This methodology has proven to be very sensible and efficient to continuous improvement processes to adapt and refine criteria for classification and possible adoption of other parameters.

Appendix

Abstract of fuzzy sets and fuzzy overlay used in the paper:



References

1. Otero I, JCN Muñoz, Hernández, M (1996) Valoración del Paisaje y del Impacto Paisajístico de las Construcciones en el Páramo Leonés. *Revista de Cartografía, SIQ Teledetección y medioambiente*. 52–74
2. Bishop I (2002) Determination of thresholds of visual impact: the case of wind turbines. *Environ Plan B-Plann Des* 29(5):707–718
3. Bishop I (2003) Assessment of visual qualities, impacts, and behaviours, in the landscape, by using measures of visibility. *Environ Plan B-Plan Des* 30(5):677–688
4. Turner et al (2001) A visibility graphs and landscape visibility analysis. *Int J Geogr Inf Sci* 15(3):221–237
5. Fisher PF (1995) An exploration of probable viewsheds in landscape planning. *Environ Plan B-Plan Des* 22(5):527–546
6. Ramos et al (1986) Visual landscape evaluation. A grid Technique. *Landsc Plan* 3:67–88
7. Polakoski K (1975) Landscape assessment of the upper great lakes basin resources: a macro-geomorphic and micro-composition analysis. *Landsc assess Values percept resources* 203–219
8. Calvo Iglesias M (2000) Análisis del paisaje con técnicas SIG
9. Escribano M, Paisaje El et al (1987) Unidades Temáticas Ambientales de la Dirección General del Medio Ambiente. MOPU, Madrid
10. MMA (2004) Guía para la Elaboración de Estudios del Medio Físico. Contenido y Metodología
11. Klement EP, Messiar R, Pap P (2000) Triangular norms. Kluwer academic publishers, Dordrecht
12. Zadeh LA (1965) Fuzzy sets. *Inf. Control* 8(3):338–353

Ship Dynamic Positioning Decoupling Control Based on ADRC

Zhengling Lei, Guo Chen and Liu Yang

Abstract In this paper, the situation of dynamic positioning ships operated in some special speed is considered, and the nonlinear ship motion model of three degrees of freedom that contains nonlinear velocity term is established. Based on active disturbance rejection theory, a kind of active disturbance rejection decoupling control method is applied in ship dynamic positioning system; compared with the closed-loop system with PID controllers, the simulation results show that ADRC possesses an advantage in decoupling and anti-interference.

Keywords Dynamic position · ADRC · Decoupling

1 Introduction

Ship dynamic positioning system (dps) is a kind of loosely coupled structure when operated at low speed, so the role of the coupling between the various degrees of freedom was usually ignored in most dynamic positioning control study. However, with the rapid development in ship and marine engineering, ship dynamic positioning system is more and more widely used in a variety of special vessels [1], and then, inevitably the dynamic positioning system will be operated at a certain special speed, in which situation the coupling phenomena between the various degrees of freedom would be shown up apparently [2]. In order to further improve the control accuracy and system stability, the research on decoupling control of dynamic positioning is of great practical significance. The strong vitality demonstrated by classical PID in today's engineering practice attracted some control researchers to re-know and re-explore the PID control mechanism. HanJingqing [3] was one of

Z. Lei (✉) · G. Chen · L. Yang

Information Science and Technology college, Dalian Maritime University, Dalian, China
e-mail: leizhengling@hotmail.com

these who put forward a method to overcome the “shortcomings” of PID and therefore proposed the active disturbance rejection control (ADRC) theory. Extended state observer is the core component of ADRC control, which can actively estimate and compensate for, in real time, the combined effects of the “internal disturbance” and “external disturbance,” forcing an otherwise unknown plant to behave like a nominal one [4]. Based on this property, the ADRC shows a strong decoupling ability [3]. The effectiveness of ADRC’s decoupling ability has been tested on several common industrial control problems, such as aircraft flight control [5], jet engine control [6, 7], chemical process control [8], induction motor power control [9], the attitude control of hypersonic vehicle [10], the ball mill pulverizing control of the power plant [11], the power control of brushless doubly fed machine [12, 13], and so on. This paper attempts to apply ADRC control in dynamic positioning control of some certain speed conditions to test its decoupling performance.

2 Ship Dynamic Positioning Motion Model

When ship dynamic positioning system was operated at some certain speed, the coupling phenomenon in the system due to nonlinear characteristics should be considered. In this situation, the ship low-frequency motion model of three degrees of freedom, in which the motions of surge, sway, and roll were considered, can be described as shown below [2, 14]:

$$M\dot{v} + C(v)v + D(v - v_c) = \tau + w \quad (1)$$

where $v = [u, v, r]^T$ denotes the LF velocity vector, $v_c = [u_c, v_c, r_c]^T$ is a vector of current velocities, τ is a vector of control forces and moments, and $w = [w_1, w_2, w_3]^T$ is a vector of zero-mean Gaussian white noise processes describing unmodeled dynamics and disturbances. Notice that r_c does not represent a physical current velocity, but can be interpreted as the effect of currents in yaw.

The nonlinear damping forces can be neglected for dynamically positioned vessels, while linear hydrodynamic damping matrix $D > 0$ and the inertia matrix including hydrodynamic added mass terms are assumed to be positive definite $M = M^T > 0$. Assuming that the starboard and the port are symmetric, M and D can be written as

$$M = \begin{pmatrix} m - X_{\dot{u}} & 0 & 0 \\ 0 & m - Y_{\dot{v}} & mx_G - Y_{\dot{r}} \\ 0 & mx_G - Y_{\dot{r}} & I_z - N_{\dot{r}} \end{pmatrix} \quad (2)$$

$$D = \begin{pmatrix} -X_u & 0 & 0 \\ 0 & -Y_v & -Y_r \\ 0 & -Y_r & -N_r \end{pmatrix} \quad (3)$$

Most of the time, the Coriolis and centrifugal matrix $C(v) = 0$ for a ship; however, it may be significant for a ship operating at some speed. $C(v)$ is a function of the elements of the inertia matrix. Note $M = \{m_{ij}\}$, $C(v)$ can be expressed as

$$C(v) = \begin{pmatrix} 0 & 0 & -m_{22}v - m_{23}r \\ 0 & 0 & m_{11}u \\ m_{22}v + m_{23}r & -m_{11}u & 0 \end{pmatrix}$$

where the non-zero elements $m_{ij} = -m_{ji}$ are defined according to (2) such that

$$\begin{aligned} m_{11} &= m - X_{\dot{u}} & m_{23} &= mx_G - Y_{\dot{r}} \\ m_{22} &= m - Y_{\dot{v}} & m_{33} &= I_z - N_{\dot{r}} \end{aligned}$$

The kinematic equation of motion for a ship is

$$\dot{\eta} = R(\psi)v \quad (4)$$

Here, $\eta = [x, y, \psi]^T$ denotes the position and orientation vector with coordinates in the earth-fixed frame, $v = [u, v, r]^T$ denotes the linear and angular velocity vector with coordinates in the body-fixed frame, and the rotation matrix $R(\psi)$ is defined as

$$R(\varphi) = \begin{pmatrix} \cos \varphi & -\sin \varphi & 0 \\ \sin \varphi & \cos \varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

There has been a lot of published control methods on ship dynamic position of low speed; however, as for the situation that a ship operating at some certain high speed, of which it not only has the characteristics of large inertia, large time delay, strong nonlinearity, and complex interference, but also strong coupling phenomenon, the research publications are comparatively few. For this situation, after reading a lot of literatures, the author found that the ADRC algorithm provides a new solution for this kind of problem.

3 The ADRC Decoupling Design Idea of Multivariable Systems

An brief introduction about ADRC how to solve the decoupling control problem of multivariable control system is given as follows [3]. Considering a system

$$\left\{ \begin{array}{l} \ddot{x}_1 = f_1(x_1, \dot{x}_1, \dots, x_m, \dot{x}_m) + b_{11}u_1 + \dots + b_{1m}u_m \\ \ddot{x}_2 = f_2(x_1, \dot{x}_1, \dots, x_m, \dot{x}_m) + b_{21}u_1 + \dots + b_{2m}u_m \\ \vdots \\ \ddot{x}_m = f_m(x_1, \dot{x}_1, \dots, x_m, \dot{x}_m) + b_{m1}u_1 + \dots + b_{mm}u_m \\ y_1 = x_1, y_2 = x_2, \dots, y_m = x_m \end{array} \right.$$

which is an m -input and m -output system, the amplification factor b_{ij} is a function $b_{ij}(x, \dot{x}, t)$ of the state variable and time.

Assume matrix

$$B(x, \dot{x}, t) = \begin{pmatrix} b_{11}(x, \dot{x}, t) & \dots & b_{1m}(x, \dot{x}, t) \\ \vdots & \ddots & \vdots \\ b_{m1}(x, \dot{x}, t) & \dots & b_{mm}(x, \dot{x}, t) \end{pmatrix}$$

is invertible. The external disturbances $f(x_1, \dot{x}_1, \dots, x_m, \dot{x}_m) = [f_1 \ f_2 \ \dots \ f_m]^T$ are considered as “dynamic coupled portion,” while $U = B(x, \dot{x}, t)u$ is considered as “static coupled portion.”

Note $x = [x_1 \ x_2 \ \dots \ x_m]^T$, $f = [f_1 \ f_2 \ \dots \ f_m]^T$, $u = [u_1 \ u_2 \ \dots \ u_m]^T$ and introduce the “virtual control” $U = B(x, \dot{x}, t)u$, the equations above can be rewritten as

$$\begin{cases} \ddot{x} = f(x, \dot{x}, t) + U \\ y = x \end{cases} \tag{5}$$

The input–output relationship of the i th loop is given as follows:

$$\begin{cases} \ddot{x}_i = f_i(x_1, \dot{x}_1, \dots, x_m, \dot{x}_m, t) + U_i \\ y_i = x_i \end{cases}$$

Thus, it has been completely decoupled between the controlled output of the i th loop and the “virtual control” U_i . Here, $f_i(x_1, \dot{x}_1, \dots, x_m, \dot{x}_m, t)$ is the external disturbance of the i th loop. Because of the ADRC’s significant characteristic of estimating and compensating for the effects of the unknown dynamics and disturbances which force an otherwise unknown plant to behave like a nominal one, y_i can surely reach the goal $y_i^*(t)$ by embedding an ADR control between U_i and y_i , as long as the target value $y_i^*(t)$ and the output y_i of the i th loop can be measured . Then, the amount of actual control can be determined by the formula $u = B^{-1}(x, \dot{x}, t)U$.

The ship dynamic positioning motion model (1) can be rewritten as follows:

$$\begin{cases} \dot{v} = f(v, v_c, w) + U \\ \dot{\eta} = R(\psi)v \end{cases} \tag{6}$$

where $f(v, v_c, w) = M^{-1}[w - C(v)v - D(v - v_c)]$, $U = M^{-1}\tau$. And it is well known that the ship's position η can be measured by the electronic positioning navigation system on the ship. So we can conclude from the above analysis that the decoupling control of dynamic positioning system can be achieved by embedding three ADRC controllers in parallel between the control vector τ and output vector η for a surface ship.

4 ADRC Design for Each Loop

In this paper, the second-order ADRC was adopted as loop controllers. Taking the longitudinal control loop as an example, the ADRC control block diagram is shown in Fig. 1 :

The algorithm of each part in this controller is as follows: [3, 15]:

(1) Tracking differentiator(TD)

$$\begin{cases} v_1(k+1) = v_1(k) + Tv_2(k) \\ v_2(k+1) = v_2(k) + Tfst(v_1(k), v_2(k), v(k), r, h) \end{cases}$$

where T indicates the sampling period, $v(k)$ marks the input signal at time k , r is a parameter which decides the track speed, while h decides filtering effect when the input signal is polluted by noise. Function fst is calculated as follows:

$$\begin{aligned} \delta &= rh, \delta_0 = \delta h, y = x_1 - u + hx_2, \\ a_0 &= \sqrt{\delta^2 + 8r|y|} \\ a &= \begin{cases} x_2 + y/h, & |y| \leq \delta_0 \\ x_2 + 0.5(a_0 - \delta)\text{sign}(y), & |y| > \delta_0 \end{cases} \\ fst &= \begin{cases} -ra/\delta, & |a| \leq \delta \\ -r\text{sign}(a), & |a| > \delta \end{cases} \end{aligned}$$

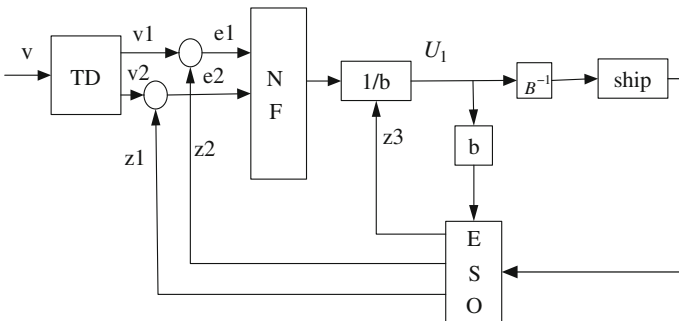


Fig. 1 Second-order ADRC controller block diagram

(2) Extended state observer (ESO equation)

$$\begin{cases} z_1(k+1) = z_1(k) + T[z_2(k) - \beta_{01}e(k)] \\ z_2(k+1) = z_2(k) \\ +T[z_3(k) - \beta_{02}\text{fal}(e(k), 1/2, \delta) + bu(k)] \\ z_3(k+1) = z_3(k) - T\beta_{03}\text{fal}(e(k), 1/4, \delta) \end{cases}$$

where $e(k) = z_1(k) - y(k)$ and

$$\text{fal}(e, a, \delta) = \begin{cases} e\delta^{a-1}, |e| \leq \delta \\ |e|^a \text{sign}(e), |e| > \delta \end{cases}$$

Like ordinary observer, the ESO accepts $u(k)$ and $y(k)$ as its input signal. Here, $u(k)$ is the coupling controller U_1 .

(3) Effects of non-smooth feedback (NF)

$$\begin{cases} e_1 = v_1(k) - z_1(k), e_2 = v_2(k) - z_2(k) \\ u_0 = \beta_1\text{fal}(e_1, a_1, \delta_1) + \beta_2\text{fal}(e_2, a_2, \delta_1) \\ u(k) = u_0 - z_3(k)/b \end{cases}$$

This controller algorithm only requires the input data $u(k)$ and output data $y(k)$ of the plant.

5 The Ship Dynamic Positioning ADR Decoupling Controller

We can draw a conclusion from the design process of the above 2,3 sections that the input of the ship dynamic positioning ADR decoupling controller based on the ship motion model can be expressed as follows (1):

$$\tau_i = \sum_{j=1}^m m_{ij}U_j \quad (7)$$

where m_{ij} represents the element of the inertia matrix and U_j is the component of the virtual control amount.

6 The Dynamic Positioning ADRC Simulation Verification

In this section, a supply ship was selected as the control plant for the study. The ship's main parameters are shown in Table 1:

Table 1 Ship's main parameters

Length overall	76.2 m	Draft	6.25 m
Beam overall	18.8 m	Net weight	4200 t
Vertical height	82.5 m	Main engine power	3533 kw

The model parameters of the supply ship were obtained through several large-scale seas by the GNC laboratory of Norwegian University of Science and Technology (NTNU), where the dimensionless inertia matrix M and damping matrix D are given as follows [14]:

$$M = \begin{pmatrix} 1.1274 & 0 & 0 \\ 0 & 1.8902 & -0.0744 \\ 0 & -0.0744 & 0.1278 \end{pmatrix} \quad (8)$$

$$D = \begin{pmatrix} 0.0358 & 0 & 0 \\ 0 & 0.1183 & -0.0124 \\ 0 & -0.0041 & 0.0308 \end{pmatrix} \quad (9)$$

The Coriolis and centrifugal matrix is

$$C(v) = \begin{pmatrix} 0 & 0 & -1.8902v + 0.0744r \\ 0 & 0 & 1.1274u \\ 1.8902v - 0.0744r & -1.1274u & 0 \end{pmatrix} \quad (10)$$

Assume that the ships' starting position is $\eta = [0, 0, 0]^T$ and the target location is $\eta = [50, 50, 10]^T$. In the case of not considering the effects of wind, wave, and current, the parameters of each ADRC controller were tuned as follows:

Longitudinal loop:

$$\begin{aligned} r &= 30, h = 0.01, T = 0.01, \beta_0 = [100, 800, 3000], \\ \beta_1 &= [2, 1.5], a = [0.75, 1.25], \\ \delta &= \delta_1 = 0.01, b = 5 \end{aligned}$$

Transverse loop:

$$\begin{aligned} r &= 30, h = 0.01, T = 0.01, \beta_0 = [100, 800, 3000], \\ \beta_1 &= [2, 1.5], a = [0.75, 1.25], \\ \delta &= \delta_1 = 0.01, b = 5 \end{aligned}$$

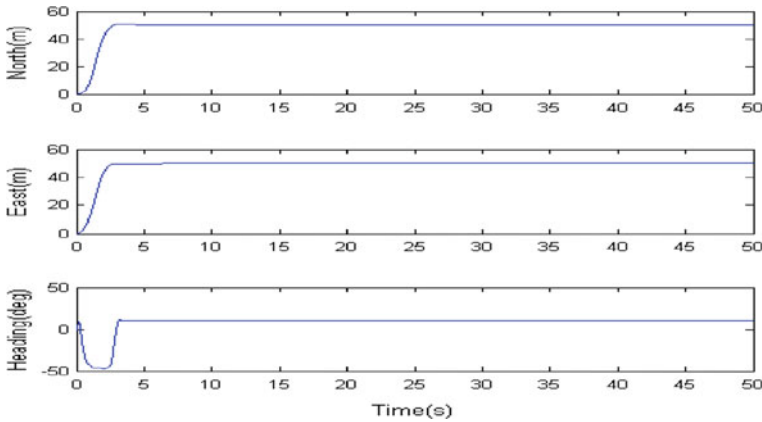


Fig. 2 Ship position response curve of ADR-CS

Yawing loop:

$$r = 30, h = 0.01, T = 0.01, \beta_0 = [200, 1000, 5000],$$

$$\beta_1 = [100, 20], a = [0.75, 1.25],$$

$$\delta = \delta_1 = 0.01, b = 1$$

System response is shown in Fig. 2:¹

It can be read from the curves that the system achieves the control objectives in 10 sec.

The parameters of each PID controller were well tuned as follows:

Longitudinal loop: $K_p = 2e - 3, K_i = 0, K_d = 0$

Transverse loop: $K_p = 1e - 2, K_i = 0, K_d = 0$

Yawing loop: $K_p = 0.5, K_i = 0, K_d = 0$

System response is as shown in Fig. 3:

It can be concluded from the curves that the system can also achieves the control objectives; however, the response speed was comparatively slow.

6.1 Decoupling Performance Test

We apply PID controllers into coupled ship dynamic positioning system and then take its simulation results as a comparison for analysis to that of decoupling control with ADR controllers. Loop controller parameters remain unchanged.

¹ To design a pretty picture layout, record “PID control system” as “PID-CS” and record “ADR control system” as “ADR-CS” in the following captions of the pictures.

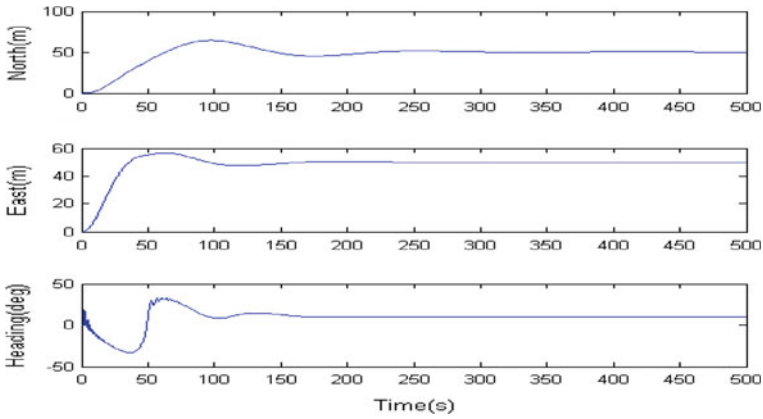


Fig. 3 Ship position response curve of PID-CS

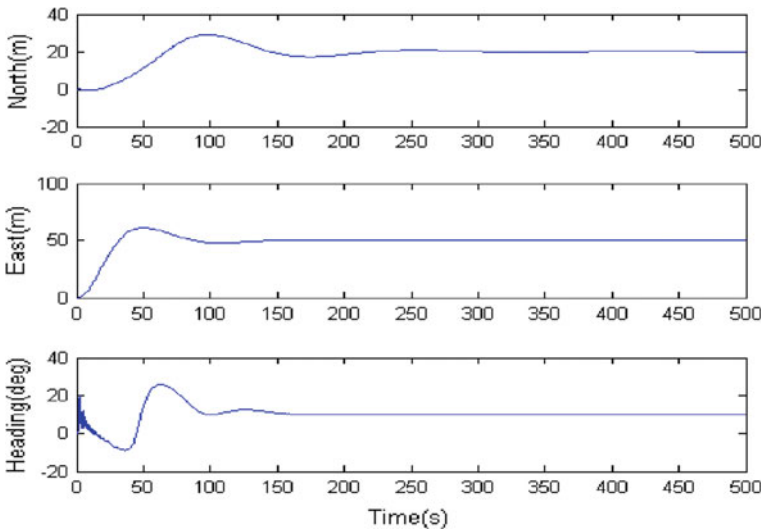


Fig. 4 Response curve of PID-CS

Assume that ship's target location is $\eta = [20, 50, 10]^T$. The response results of the two control systems are shown in Figs. 4 and 5:

The simulation results show that both the PID control system and the ADRC control system are able to achieve the goal. The reason is that the ship dynamic positioning system is a loosely coupled structure, and the coupling relationship between surge and other two degrees of freedom is the most loosest one in particular.

Assuming ship's target location as $\eta = [20, 40-50]^T$, the response results of the two control systems are shown in Figs. 6 and 7:

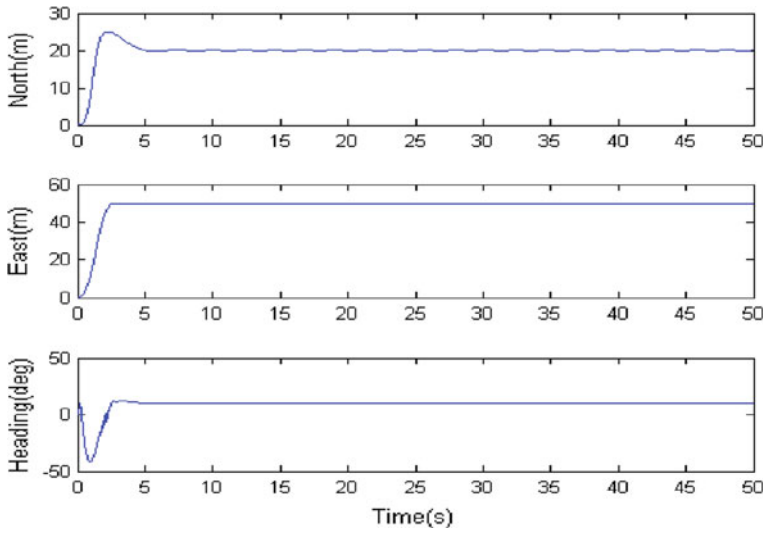


Fig. 5 Response curve of ADR-CS

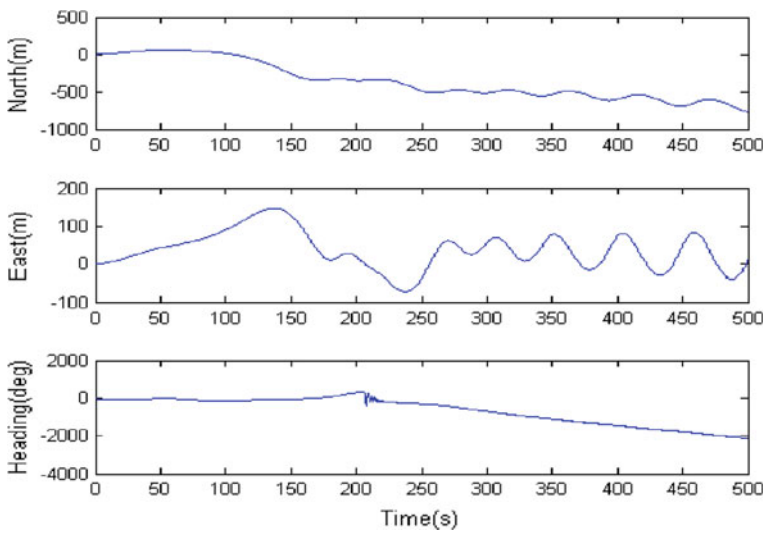


Fig. 6 Response curve of PID-CS

It can be read from the curves that the PID control system has been completely out of position, while the ADR control system can still point target very well, which confirms that the ADR controller has strong decoupling performance.

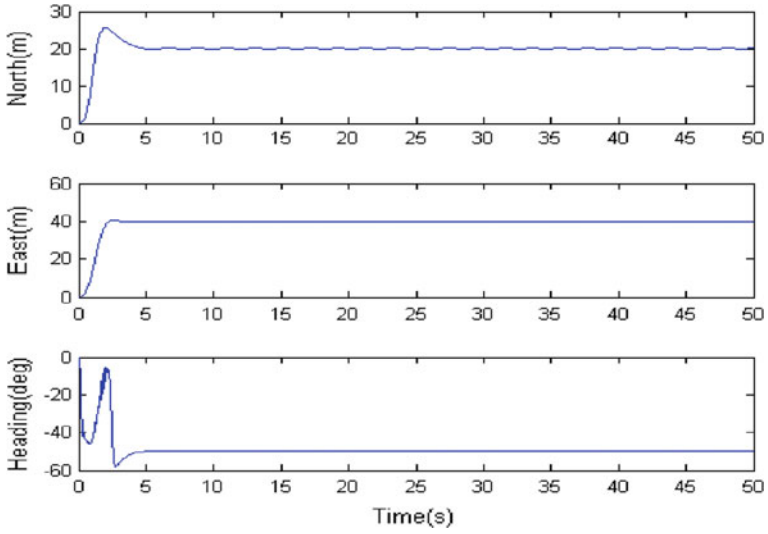


Fig. 7 Response curve of ADR-CS

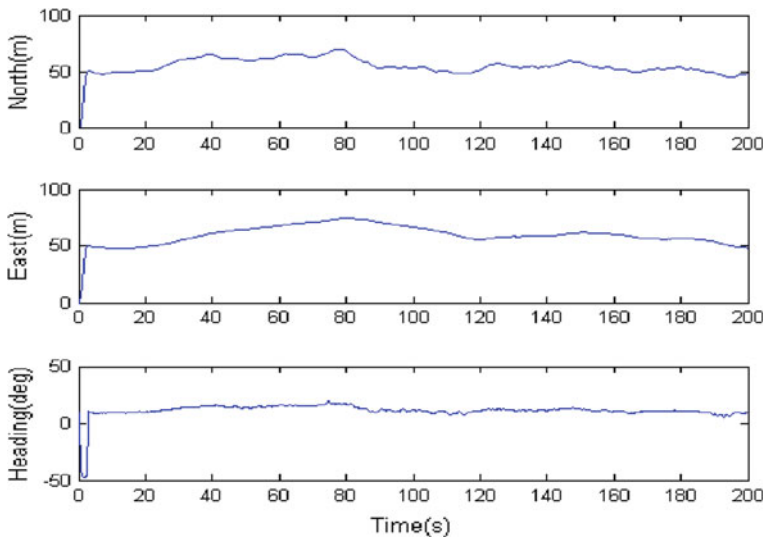


Fig. 8 Response curve of ADR-CS

6.2 System Anti-Interference Performance Test

The unmodeled dynamics and disturbances of the system can be described as follows [2, 16]:

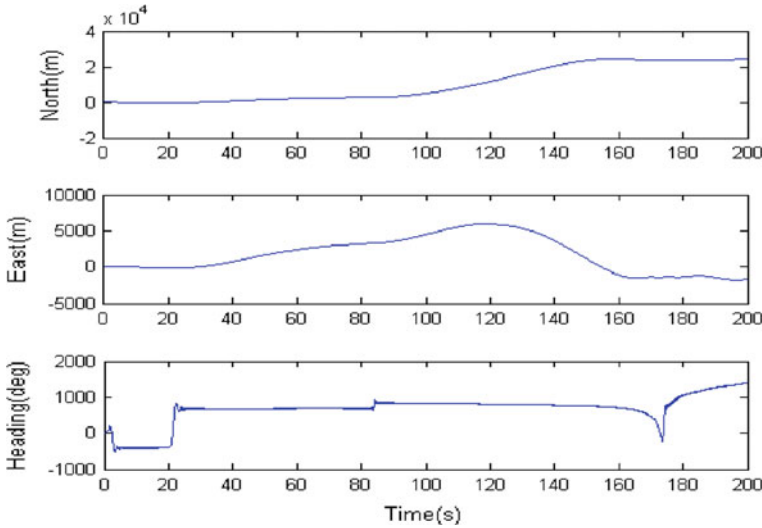


Fig. 9 Response curve of PID-CS

$$\begin{cases} \dot{w} = J^T(\eta)b \\ \dot{b} = -T_b^{-1}b + E_b\omega_b \end{cases} \quad (11)$$

Here, $E_b = \text{diag}\{E_{b1}, E_{b2}, E_{b3}\}$, ω_b represents zero-mean Gaussian white noise, and T_b is a diagonal matrix of positive bias time constants. Assume that the energy amplitude of the white noise is 1, $E_b = \text{diag}\{2, 2, 2\}$ and $T_b = \text{diag}\{500, 500, 500\}$. The response results of the two control systems are shown in Figs. 8 and 9:

These curves above show that the ADR control system can basically control the bow in the target direction in the case of interference. Despite an oscillation error, the ship’s position was still controlled around the target location, while the PID control system has been completely out of position under the same situation.

7 Conclusion

The ADR control technique’s decoupling performance and anti-interference performance in ship dynamic positioning system are tested. It is shown that ADRC possesses a significant advantage in decoupling and anti-interference. The verification of ADRC’s effectiveness in ship dynamic positioning control further confirms its potential as a transformative control technology. Even so excellent it behaves, however, the ADR controller based on nonlinear extended state observer encounters a big challenge in the aspect of parameter tuning, which makes it not easy to be popularized. So the parameter tuning techniques of nonlinear ADR control will be a worth exploring direction.

Acknowledgments This work was supported by The National Natural Science Foundation of China (No. 61074053) and The Applied Basic Research Program of Ministry of Transport of China (No. 2011-329-225-390).

References

1. Mogen MJ (1985) Dynamic position of offshore ships. National Defence Industry Press, Beijing
2. Fossen TI (1994) Guidance and control of ocean vehicles. Wiley, New York.
3. H. J.Q (2008) The ADRC control technology: estimated compensate for uncertainties control technology. National Defence Industry Press
4. Zheng Q, Gao Z (2010) On practical applications of active disturbance rejection control. In: Proceedings of the 2010 Chinese Control Conference
5. Huang Y, Xu K, Han J, Lam J (2001) Flight control design using extended state observer and non-smooth feedback. In: Decision and control, Proceedings of the 40th IEEE Conference on, vol. 1. IEEE 2001:223–228
6. Miklosovic, R, Gao Z (2005) A dynamic decoupling method for controlling high performance turbofan engines. In: Proceeding of the 16th IFAC World Congress, pp 4–8
7. Zhang HB, Wang JK, Wang RX, Sun JG (2012) Design of an active disturbance rejection decoupling multivariable control scheme for aero-engine. *J Propul Techno* 1:78–83
8. Zheng Q, Chen Z, Gao (2007) A dynamic decoupling control approach and its applications to chemical processes. American Control Conference, 2007. ACC '07 July 2007, pp. 5176–5181
9. Liu J, Huang L, Kang ZJ (2012) Decoupling control of power in double-fed induction generator based on auto-disturbance rejection control technology. *Electric Mach Control Appl* 1:57–61
10. Qi NM, Qin CM, Song ZG (2011) Improved ADRC cascade decoupling controller design of hypersonic vehicle. *J Harbin Inst Techno* 11:34–38
11. M YG, H N, L PF, L Y (2007) ADRC-based multivariate decoupling control of the ball mill applications. *J Eng Thermal Energy Power* 3:297–300+347
12. Z XY, S J, Y JM (2007) Power control strategy based on auto -disturbance rejection decoupling for a variable speed constant frequency generation system. *Electr Drive* 2:8–11+35
13. Z XY, S J, W J (2008) ADRC power decoupling control of brushless doubly-fed wind turbine. *Acta Energiæ Solaris Sinica* 12:1477–1483
14. Fossen T, Sagatun S, Sørensen A (1996) Identification of dynamically positioned ships. *Control Eng Practice* 4(3):369–376
15. Xue DY, Chen YQ (2002) System simulation technology and application. Tsinghua university press, Beijing
16. Fossen T, Strand J (1999) Passive nonlinear observer design for ships using Lyapunov methods: full-scale experiments with a supply vessel. *Automatica Oxford* 35:3–16

Intelligent Double-Eccentric Disc Normal Adjustment Cell in Robotic Drilling

Peijiang Yuan, Maozhen Gong, Tianmiao Wang, Fucun Ma,
Qishen Wang, Jian Guo and Dongdong Chen

Abstract An intelligent verticality adjustment method named double-eccentric disc normal adjustment (DENA) is presented in precise robotic drilling for aero-structures. The DENA concept is conceived specifically to address the deviation of the spindle from the surface normal at the drilling point. Following the concept of intelligent and accurate normal adjustment, two precise eccentric discs (PEDs) with the identical eccentric radius are adopted. Indispensably, two high-resolution stepper motors are used to provide rotational power for the two PEDs. Once driven to rotate with appropriate angles respectively, two PEDs will carry the spindle to coincide with the surface normal, keeping the vertex of the drill bit still to avoid the repeated adjustment with the help of the spherical plain bearing. Since the center of the spherical plain bearing coincides with the vertex of the drill bit, successful implementation of DENA has been accomplished on an aeronautical drilling robot platform. The experimental results validate that DENA in robotic drilling is attainable in terms of intelligence and accuracy.

Keywords Double-eccentric disc · Normal adjustment · Robotic drilling · Aero-structures

P. Yuan (✉) · M. Gong · T. Wang · F. Ma · Q. Wang · D. Chen
School of Mechanical Engineering and Automation, Beihang University,
100191 Beijing, China
e-mail: itr@buaa.edu.cn

M. Gong
e-mail: gongmaozhen@126.com

T. Wang
e-mail: itm@buaa.edu.cn

F. Ma
e-mail: giantmfc@163.com

J. Guo
Xinanhe Waste Water Treatment Plant, Municipal Drainage Administration,
264003 Yantai, China

1 Introduction

Once limited by the vertical accuracy of the hole drilled in aero-structures skin, there will be a significant drop in the drilling quality. The oblique hole can give rise to severe fatigue cracks, which will threaten the safety and fatigue life of the aircraft [1, 2]. In modern aircraft design, many titanium alloy [3], super alloys [4], and composite structures [5] are adopted to enhance aircraft structure strength, which highly improve the fatigue life and reduce the weight of the aircraft. However, when these difficult-to-cut metals are drilled in a deflective direction from the surface normal at the drilling point, the drilling force will greatly increase, which makes it more difficult to be drilled. For this reason, the normal adjustment before drilling is highly desired.

With the development of robot technology, the aircraft digital assembly has been interested in the use of robotic drilling [6]. Representatively, the robotic drilling system based on industrial articulated arm robot is widely used in drilling for aero-structures [7]. ElectroImpact cooperated with Boeing has developed the flex-track drilling robot [8]. In addition, the crawling drilling robot has also become a hot spot of research [9]. The traditional method to guide the drilling in surface normal direction is to use a drilling template [10]. But robotic drilling is well known for the accuracy and efficiency, so the drilling template cannot meet its requirement, so that it is urgent to make intensive research in automatic normal adjustment. At present, there are mainly two approaches of normal adjustment: one is that both the spindle apparatus and workpiece are adjusted each normal attitude; the other is that only spindle apparatus realizes attitude adjustment, keeping workpiece still [11]. The former mainly adjusts the processed workpiece to accomplish the relative attitude adjustment of spindle apparatus, such as three-point bracket regulation algorithm [12] and the drilling equipment with automatic angle adjustment [13], which is difficult to meet the processing-site demand of drilling for large workpiece and real-time normal adjustment. The four or five axes drilling apparatus are used to realize vertical fine adjustment of the spindle [14]. But it is not only expensive but large in size. Besides, the verticality is adjusted by joint motions [15, 16], which is small in terms of volumes, but long adjustment path. In order to address the above-noted problems, an intelligent verticality adjustment method named double-eccentric disc normal adjustment (DENA) is presented in precise robotic drilling for aero-structures.

2 DENA Cell

The DENA cell mounted on the end effector of the aero-drilling robot provides a problem-solving platform to make the aero-structures drilled along the direction of the surface normal at the drilling point. As an automated and high-precision adjustment cell, the automatic measurement and detection of the surface normal at

the drilling point are also indispensable. Once the surface normal measured using “4-point tangent-cross product method” in [17], the unit surface normal vector will be fed back to the controller, as shown in Fig. 1. According to the feedback, the controller will make a decision to drive two high-resolution stepper motors synchronously, which can cause two precise eccentric discs (PEDs) (the small one and the big one) to rotate with appropriate angles, respectively, in form of gear drive. Owing to the eccentricity in different position, the spindle inset the small PED is carried to anywhere within the adjustment area of $\pm 5^\circ$ result from two PEDs’ resultant motion.

Ultimately, the axis of the spindle coincides with the surface normal at the drilling point. While the spindle adjusted, the vertex of the drill bit keeps still, avoiding the position compensation and repetitive positioning of drilling tool. Meanwhile, the deviation between the surface normal vector and the axis of the spindle needs to be detected. If the deviation is within 0.5° , the robot will execute the drilling task. Otherwise, the computer will issue instructions to repeat the above-mentioned normal adjustment process until the deviation between the surface normal vector and the axis of the drill bit is less than 0.5° . Delightedly, the vertical accuracy has successfully been placed within the specification’s $\pm 0.1^\circ$ tolerance.

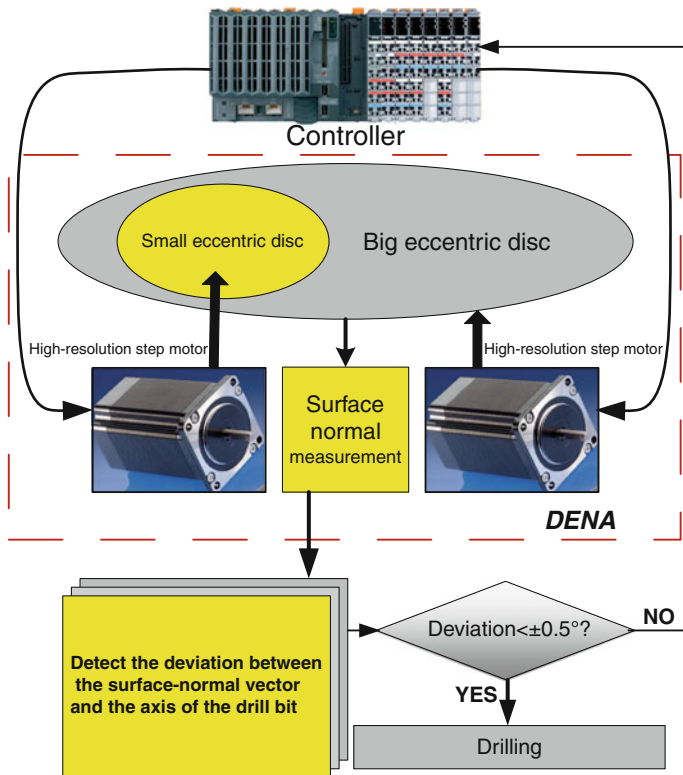


Fig. 1 Flowchart of the normal adjustment process of DENA cell

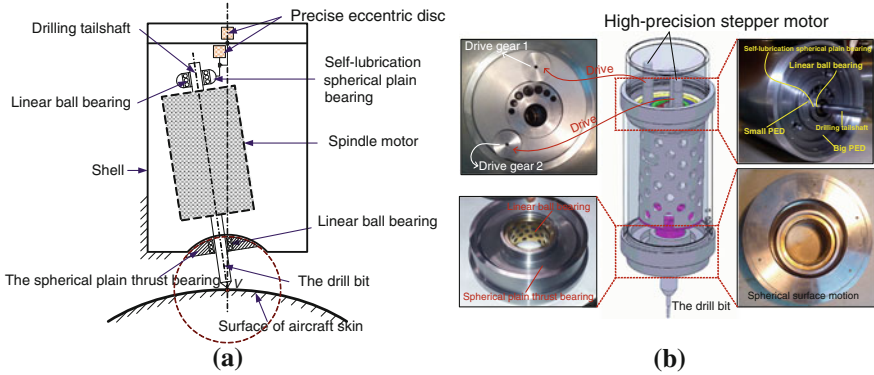


Fig. 2 Structural and schematic diagram of working principle of DENA cell

3 Innovative Design of DENA

DENA provides a precise mechanical actuator with 2-DOF for automated and precise robotic drilling, which is aimed to accomplish the fine normal adjustment of the spindle. In order to fulfill the accurate and vertical adjustment of the spindle, DENA is designed to adopt the servo circular motion mechanism with two PEDs as shown in Fig. 2a, which make two rotating DOFs (R_x and R_y) along the axial direction adjusted finely, meanwhile, provide the axial feed for drilling. The end effector with DENA is mounted on a flex-track robot. DENA, spindle motor, drill clamp, and drill bit situate on a spherical plain thrust bearing, whose center coincides with the vertex of the drill bit to keep it still when adjusting, avoiding the repeated adjustment and position compensation. The drill bit connected with the spindle motor by a coupling and drill clamp is mounted on the shell of the end effector through the linear ball bearing and spherical plain thrust bearing. The axis of the spindle can freely rotate in a coning motion around the axis of end effector, keeping the point V still. In addition, the linear ball bearing also provides an oriented track for the spindle feed during drilling process.

Self-lubrication spherical plain bearing provides a solid foundation in which the linear ball bearing is mounted. The drilling tailshaft at the rear end of spindle motor, which is coaxial with the axis of the drill bit, passes through the linear ball bearing. There are two PEDs with the identical eccentric radius r . The small one, in which self-lubrication spherical plain bearing situates, is embedded in the big one and can rotate freely. In this way, when these two PEDs rotate, respectively, to an appropriate angle from the initial position, the normal adjustment of the spindle can be accomplished.

In order to adjust the verticality of the spindle automatically and control easily, two high-resolution stepper motors with a constant step angle of 1.8° are used to drive two PEDs by two drive gear, as shown in Fig. 2b. The stepper motor driver adopted in this paper is fractionized evenly into 1,000. In other words, when a

pulse is emitted, the stepper will rotate 0.0018° . The standard pitch diameters of the gear attached to the big PED and the small one are 200 and 100 mm, respectively, and the standard pitch diameters of two drive gears connected with the high-resolution stepper motors are 50 mm. So, the transmission ratio between PEDs and two drive gears is 4:1 and 2:1. When a pulse produced, the big PED will rotate $0.0018/4^\circ$ and the small one will rotate $0.0018/2^\circ$, which not only guarantees the precision of the normal adjustment but also meets the requirement of automatic adjustment.

4 Adjustment Algorithm of DENA Cell

As Fig. 3a shows, the distance between the vertex V and the adjustment plane π of two PEDs is expressed as D . Based on the surface normal vector at the drilling point, the normal adjustment position O_2 at the plane π can be obtained. As is described in Fig. 3a, the coordinate system is established. We suppose the unit surface normal vector n , as $n = (x_n, y_n, z_n)$. The straight line L which passes through the point O and takes the vector n as its direction vector can be expressed as follows

$$\frac{x - 0}{x_n} = \frac{y - 0}{y_n} = \frac{z - 0}{z_n} \tag{1}$$

Since there is an intersection O_2 between the straight line L and the plane π , we can obtain the coordinate of point O_2 .

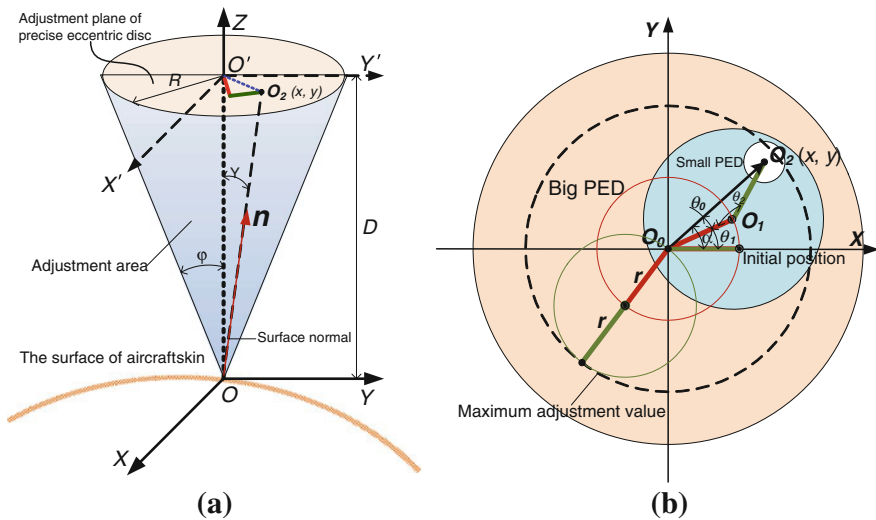


Fig. 3 Fine adjustment of DENA cell through the rotation of two PEDs

$$(x_{O_2}, y_{O_2}, z_{O_2}) = \left(\frac{Dx_n}{z_n}, \frac{Dy_n}{z_n}, D \right) \quad (2)$$

where D represents the distance between the vertex of the drill bit and the adjustment plane π of two PEDs.

Meanwhile, the included angle γ between the unit surface normal vector \mathbf{n} and the axis of the spindle can be expressed as:

$$\gamma = \cos^{-1}[(x_n, y_n, z_n) \cdot (0, 0, 1)] \quad (3)$$

Two PEDs have the identical eccentric radius r , as shown in Fig. 3b. The small PED is mounted in the eccentric circle of the big one, around whose center the small PED can rotate freely. The drilling tailshaft of the spindle passes through the eccentric circle of the small PED. When adjusting, the spindle will be carried to the normal adjustment position O_2 . In the initial state, the point O_2 coincides with the center of the big PED. However, when the point O_0 , O_1 , and O_2 are collinear, the maximum adjustment distance is reached, namely the radius $R = 2r$ in Fig. 3b. In DENA cell, the distance D is 400 mm, so the maximum adjustment angle can be obtained.

$$\varphi = \tan^{-1}\left(\frac{2r}{D}\right) = \tan^{-1}\left(\frac{2 \times 17.5}{400}\right) = 5^\circ \quad (4)$$

where $r = 17.5$ mm. Hence, the maximum adjustment area of DENA is $\pm 5^\circ$.

To know how many angles two PEDs rotate, just figure out the angle θ_1 and θ_2 . Without loss of generality, suppose that the point O_2 in Fig. 3b is the normal adjustment position. So, the distance between the point O_0 and O_1 can be written as

$$d_{O_0O_2} = \sqrt{x_{O_2}^2 + y_{O_2}^2} \quad (5)$$

Likewise, the angle α can be expressed as

$$\alpha = \begin{cases} \cos^{-1} \frac{(1,0) \cdot (x_{O_2}, y_{O_2})}{\|(x_{O_2}, y_{O_2})\|}, & \text{If } y \geq 0 \\ 2\pi - \cos^{-1} \frac{(1,0) \cdot (x_{O_2}, y_{O_2})}{\|(x_{O_2}, y_{O_2})\|}, & \text{If } y < 0 \end{cases} \quad (6)$$

where the angle α will range from 0 to 360° .

Obviously, $\Delta O_0O_1O_2$ is an isosceles triangle in Fig. 3b. According to the geometric property of the isosceles triangle, the angles both θ_0 and θ_2 are available.

$$\begin{cases} \theta_0 = \cos^{-1} \frac{d_{O_0O_2}}{2r} = \cos^{-1} \frac{\sqrt{x_{O_2}^2 + y_{O_2}^2}}{2r} \\ \theta_2 = \pi - 2\theta_0 \end{cases} \quad (7)$$

According to Eqs. (6) and (7), we can get

$$\begin{cases} \theta_1 = \alpha - \theta_0 = \alpha - \cos^{-1} \frac{\sqrt{x_{O_2}^2 + y_{O_2}^2}}{2r} \\ \theta_2 = \pi - 2 \cos^{-1} \frac{\sqrt{x_{O_2}^2 + y_{O_2}^2}}{2r} \end{cases} \quad (8)$$

So that, we can figure out the number of pulses which two high-resolution stepper motors need. They can be expressed as follows:

$$N_{\text{pulse1}} = \frac{\theta_1 \cdot N \cdot R_{B-PED}}{\Delta\theta \cdot R_{DG}}, \quad N_{\text{pulse2}} = \frac{\theta_2 \cdot N \cdot R_{M-PED}}{\Delta\theta \cdot R_{DG}} \quad (9)$$

where

- N_{pulse1} Number of pulses sent by motor 1;
- N_{pulse2} Number of pulses sent by motor 2;
- N Subdivision number of stepper motors;
- $RB-PED$ Standard pitch radius of big PED;
- $RM-PED$ Standard pitch radius of small PED;
- RDG Radius of the drive gear;
- $\Delta\theta$ Step angle of stepper motors.

As long as the number (N_{pulse1} and N_{pulse2}) of pulses sent by two high-resolution stepper motors are obtained, two PEDs are driven to rotate respective certain angle as mentioned previously (θ_1 and θ_2), the spindle can be adjusted to drill along the surface normal direction at the drilling point. However, there are four different schemes to realize the normal adjustment as follows:

1. Remaining the big PED still, rotate the small one with θ_2 degrees. Once the small PED finishes, the big one will rotate θ_1 degrees immediately. For more explicit description, a movement simulation is given. Consequently, the movement locus of the center O_1 and O_2 of two PEDs is available. Take O_2 (20, 25) for example, as shown in Fig. 4a.
2. Contrary to what is stated, remaining the small PED still, rotate the big one with θ_1 degrees. Once the big PED finishes, the small one will rotate θ_2 degrees immediately as shown in Fig. 4b.
3. Simultaneously, both the two PEDs are started to rotate respective angles (θ_1 and θ_2) at the same speed. During the adjustment process, the big PED stops first, then the small one stops, as shown in Fig. 4c.

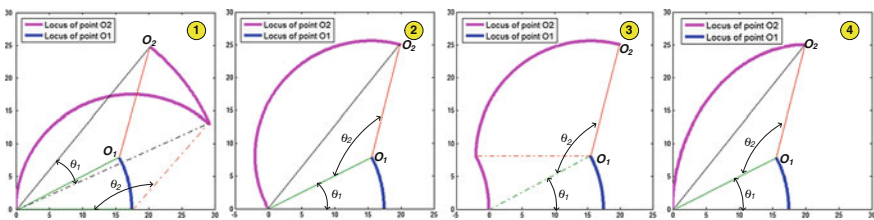


Fig. 4 Movement locus of the center O_1 and O_2 of two PEDs

4. Within a same time, both the two PEDs begin and stop at the same time according to following Eq.

$$\begin{cases} x = r[\cos \theta_1 - \cos(\theta_1 - \theta_2)] \\ y = r[\sin \theta_1 - \sin(\theta_1 - \theta_2)] \end{cases} \quad (11)$$

The movement locus is shown in Fig. 4d.

Comparing these four schemes, it is pretty clear that the fourth scheme provides a smoother normal adjustment along the shortest path. A complete normal adjustment process can be summarized as follows: Based on the unit surface normal vector, the normal adjustment position is figured out, followed by the calculation of the θ_1 and θ_2 . Finally, the controller makes the two high-resolution stepper motors to drive the PEDs which carry the spindle to coincide with the surface normal at the drilling point.

5 Experiments and Analysis

A drilling experiment, which takes normal adjustment of the spindle into account, is conducted on a flex-track drilling robot platform, as shown in Fig. 5. Throughout the experiment, there are two parts including the normal adjustment and drilling for aluminum alloy (Al), carbon fiber reinforced plastics (CFRP), and titanium alloy (Ti) aero-structures. The effects of normal adjustment on drilling quality, drilling force, and thermal damage are also analyzed.

Fig. 5 Flex-track drilling robot experiment platform with DENA cell

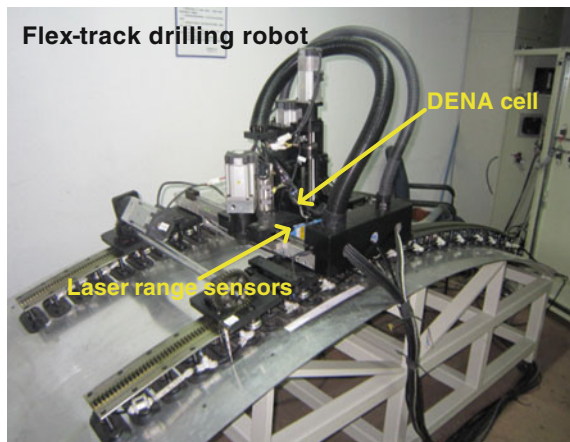


Table 1 Experimental data of normal adjustment

No.	Surface normal vector n	γ (degrees)	O_2	
1	(0.049841, 0.062301, 0.996812)	4.576276	(20.000161, 25.000100)	
2	(-0.011084, 0.014051, 0.999840)	1.025482	(-4.434309, 5.621041)	
3	(-0.013700, 0.006398, 0.999886)	0.866346	(-5.480625, 2.559492)	
No.	θ_1 (degrees)	θ_2 (degrees)	N_{pulse1}	N_{pulse2}
1	27.500012	132.335403	61,111	147,039
2	50.072754	23.607329	111,273	26,230
3	74.919105	19.904047	166,487	22,116

5.1 Normal Adjustment

Based on the unit surface normal vector, three sets of data for Al, Ti, and CFRP are figured out, as listed in Table 1. Then, controller will issue commands to drive two high-resolution stepper motors which will make two PEDs rotate respective angles. As a result, the spindle will realize 3D coning motion. Ultimately, the axis of the spindle coincides with the surface normal, keeping the vertex of the drill bit still and avoiding the repeated adjustment.

5.2 Drilling and Drilling Quality

The high-quality and automated drilling is our ultimate purpose of designing the DENA cell. So, the precision and automaticity of DENA cell must be confirmed by large drilling experiment. Al, Ti, and CFRP are selected as the drilling materials which have been drilled by the drilling robot with DENA cell.

1. *Drilling for Al:* A carbide drill and countersink tool are used to drill for Al with the thickness of 6 mm. The diameter of the drilling part is 6 mm, and the taper angle of the countersinking part is 100°. What is more, the countersinking depth is 2 mm. When drilling, the spindle speed is set at 2,400 rpm/min and the feeding speed is 0.2 mm/r. The countersinking is with a spindle speed of 1,200 r/min and a feed speed of 0.12 mm/r, the drilling effect is shown in Fig. 6 and Table 2.

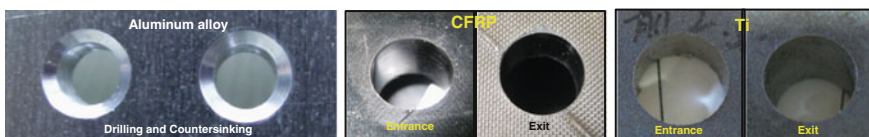


Fig. 6 Drilling effect using DENA cell for Al, CFRP, and Ti

Table 2 Robotic drilling indexes using DENA cell for Al, CFRP, and Ti

Material	Bore diameter (mm)	Countersinking angle (degrees)	Surface roughness (Ra)	Exit burr height (μm)	Delamination factor
Al	6.013–6.020	99.28–100.32	1.0–1.1	15–30	–
CFRP	12.005–12.012	–	0.8–1.0	–	1.0183–1.0508
Ti	12.010–12.021	–	0.9–1.1	50–80	–

2. *Drilling for CFRP*: The thickness of CFRP is 8 mm, and the ultimate diameter is 12 mm. We give two processes (drilling and reaming): drilling ($R1 = 9.5$ mm) \rightarrow reaming1 \rightarrow ($R2 = 11.557$ mm) \rightarrow reaming2($R3 = 12$ mm). The rotation speed of spindle is set as 6,000 rpm for CFRP. When reaming, the rotation speed of spindle and the feed speed are 1,000 rpm and 25 mm/min.
3. *Drilling for Ti*: The thickness of Ti is 8 mm, and the ultimate diameter is also 12 mm. We give two same processes as CFRP: drilling ($R1 = 9.5$ mm) \rightarrow reaming1 \rightarrow ($R2 = 11.557$ mm) \rightarrow reaming2($R3 = 12$ mm). The rotation speed of spindle is set as 800 rpm, and the feed speed is 10 mm/min. The rotation speed of spindle and the feed speed during reaming are also 1,000 rpm and 25 mm/min, respectively.

5.3 Effect of DENA Cell on Bore Diameter

The deviation of the spindle can result in the elliptical hole, which will impact on aircraft assembly. Table 2 shows the diameters of the holes drilled by robotic drilling with DENA. The maximal bore diameter is 12.012 mm and the minimal one is 12.005 mm for CFRP, so the tolerance zone has only 0.007 mm. For Ti, the maximal bore diameter is 12.021 mm and the minimal one is 12.010 mm, and the tolerance zone has 0.011 mm. All are within “ $\Phi 12H7$,” which improve the circularity of the drilling.

5.4 Effect of DENA Cell on Drilling Force

The deviation of the spindle will tremendously impact on the drilling force as shown in Fig. 7. What shows is drilling force during the drilling for Al, whose thickness is

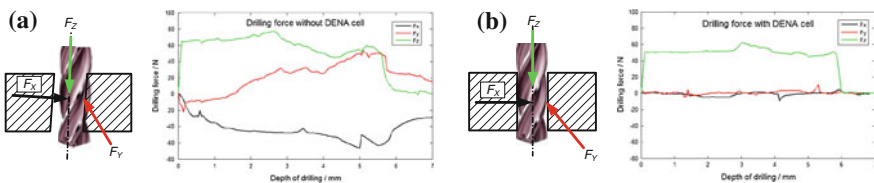


Fig. 7 Drilling force in ordinary drilling and robotic drilling with DENA cell

6 mm, with DENA cell and without DENA cell. Due to the deviation of the spindle, both the radial force F_x and the tangential force F_y increase drastically, and $F_{x_{\max}} = 67 \text{ N}$, $F_{y_{\max}} = 54 \text{ N}$. The robotic drilling with DENA makes that: $F_{x_{\max}} = 9 \text{ N}$ and $F_{y_{\max}} = 11 \text{ N}$. Likewise, the maximal axial force F_z reduces to 62 N from 77 N. Beyond that, the normal adjustment of the spindle using DENA cell makes the drill process, ranging from 0 to 6 mm, become more smoothly. Furthermore, the retracting cutter process, which has been subjected to the resistance caused by the radial force F_x and the tangential force F_y , becomes more easy now.

6 Conclusions

In this paper, the DENA concept is conceived specifically to address the deviation of the spindle from the surface normal at the drilling point. Once driven to rotate with appropriate angles, respectively, by two high-resolution stepper motors, two PEDs will carry the spindle to coincide with surface normal. Since the vertex of the drill bit coincides with the center of the spherical plain thrust bearing, the axis of the spindle can freely rotate in a coning motion around the axis of end effector, keeping the vertex of the drill bit still and avoiding the repeated adjustment and position compensation. During the whole adjustment process, two high-resolution stepper motors with a constant step angle of 1.8° are used to drive two PEDs by the gear drive, which meets the requirements of automatic adjustment. Based on the surface normal vector, DENA cell is responsible for the verticality adjustment of the spindle, which can realize the adjustment of $\pm 5^\circ$ around the axis of the spindle. The adjusted vertical accuracy lies within $\pm 0.1^\circ$. The experiment conducted on a flex-track drilling robot validates that DENA cell is attainable in terms of intelligence and accuracy. All experimental analyses indicate that policies presented in this paper outperform the comparative ones in the ordinary robotic drilling.

Acknowledgments This work is partially supported by National High Technology Research and Development Program (863 Program) of China under grant No. 2011AA040902, National Natural Science Foundation of China under grant No. 61075084, and Fund of National Engineering and Research Center for Commercial Aircraft Manufacturing, (Project No. is SAmc12-7s-15-020).

References

1. Molent L (2010) Fatigue crack growth from flaws in combat aircraft. *J Fat* 32:639–649
2. Gobbato M, Conte JP, Kosmatka JB, Farrar CR (2012) A reliability-based framework for fatigue damage prognosis of composite aircraft structures. *Probab Eng Mech* 29:176–188
3. Bi SS, Liang J (2011) Robotic drilling system for titanium structures. *Int J Adv Manuf Technol* 54:767–774
4. Henderson AJ, Bunget C, Kurfess TR (2010) Cutting force modeling when milling nickel-base superalloys. In: ASME international manufacturing science and engineering conference (MSEC), Pennsylvania, USA, pp 193–202

5. Tsao CC (2008) Experimental study of drilling composite materials with step-core drill. *Mater Des* 29:1740–1744
6. Woodruff N (2007) Airbus aims high [Robotics Drilling]. *Manuf. IET* 86:26–31
7. Atkinson J, Hartmann J, Jones S, Gleeson P (2007) Robotic drilling system for 737 aileron. In: SAE aerospace technology congress and exhibition 2007-01-3821. Los Angeles, CA, USA
8. Thompson P, Hartmann J, Feikert E, Buttrick J (2005) Flex track for use in production. *SAE Trans* 114:1039–1045
9. White TS, Alexander R, Callow G, Cooke A, Harris S, Sargent J (2005) A mobile climbing robot for high precision manufacture and inspection of aerostructures. *Int J Robot Res* 24:589–598
10. Day A, Stanley BD (2005) Apparatus and method for drilling holes and optionally inserting fasteners. US Patent 6905291
11. Shan YC, He N, Li L, Yang YF (2011) Realization of spindle prompt normal posture alignment for assembly holemaking on large suspended panel. In: International conference measuring technology and mechatronics automation. Shanghai, China, pp 950–956
12. Qin XS, Wang WD, Lou AL (2007) Three-point bracket regulation algorithm for drilling and riveting of aerofoil. *Acta Aeronautica et Astronautica Sinica* 28:1455–1460
13. Xia CH (2011) Drilling equipment with automatic angle-adjustment. Patent 201833210U
14. Speller TH Sr, Davern JW (1993) Five axis riveter and system. US Patent 5248074
15. Marguet B, Wiegert F, Lebahar O, Bretagnol B, Okcu F, Ingvar E (2007) Advanced portable orbital-drilling unit for airbus final assembly lines. In: SAE aerospace technology conference and exposition. 2007-01-3849. Los Angeles, CA, USA
16. Yamazaki K, Tomono M, Tsubouchi T (2008) Pose planning for a mobile manipulator based on joint motions for posture adjustment to end-effector error. *Adv Rob* 22:411–431
17. Gong MZ, Yuan PJ, Wang TM, Yu LB (2012) A novel method of surface-normal measurement in robotic drilling for aircraft fuselage using three laser range sensors. In: International conference on advanced intelligent mechatronics (AIM), Kaohsiung, Taiwan, pp 450–455

A Method of Fuzzy Attribute Reduction Based on Covering Information System

Fachao Li, Jiaoying Wang and Chenxia Jin

Abstract In a covering decision system, the outside edge of attributes is not clear and the existing attribute reduction methods have limitations. In this paper, we first put forward the idea of fuzzy attribute; second, we determine the cover through a δ value, and further, we can give a method of attribute reduction by using the information entropy; finally, we compare our method with [15] through a practical case and prove that our reduction method is more practical than that of [15], and attribute reduction results vary with the standard of covering.

Keywords Fuzzy attribute · Covering · Attribute reduction · Information entropy

1 Introduction

The theory of rough sets, a method of dealing with the uncertainty by deterministic method, was proposed by Pawlak [1]. It can obtain knowledge from the data or experience. Now, rough set theory has been successfully applied to the fields of machine learning, decision analysis, process control, pattern recognition, and knowledge discovery in databases. But it is hard to avoid the missing or noise interference of data in the practical problems. Pawlak's rough sets model could not handle missing value problem. Combining with different background, many scholars studied the promotion of Pawlak's rough sets model. For example, Ziarko [2]

F. Li (✉) · C. Jin

School of Economy and Management, Hebei University of Science and Technology,
Shijiazhuang 050018, China
e-mail: lifachao@tsinghua.org.cn

F. Li · J. Wang

School of Science, Hebei University of Science and Technology,
Shijiazhuang 050018, China
e-mail: wjiaoying@126.com

put forward the variable precision rough set model, which allowed a certain degree of misclassification rate. This model is conducive to find the potential law from the seemingly unrelated data. Dubois [3, 4] creatively combined rough set theory and fuzzy set theory and proposed the concept of rough fuzzy sets and fuzzy rough sets in 1900. The axiomatic of fuzzy rough sets has been studied by Morsi [5], but the research was limited to the fuzzy similarity relation. The structure and axiomatic of the dual fuzzy rough approximate operator under general fuzzy relations have been discussed by Wu [6] who used the minimum justice set to describe the dual fuzzy rough approximate operator of various fuzzy relations. Mi [7] delineated the axiomatic of T-fuzzy rough approximate operator and measured the uncertainty of T-fuzzy rough sets using the principle of comentropy. These theories provided a theoretical basis to deal with more complex problems.

Traditional rough set theory is based on indiscernible relation. There is no intersection between the concepts in knowledge. But it is difficult to get the accurate partition in many practical problems. So, the application of Pawlak's rough set model is limited in many areas. Therefore, many scholars worked on modifications of Pawlak's rough sets model to overcome such a problem. For example, Zakowski [8–11] proposed covering rough sets model and discussed the related properties. In 2003, Zhu and Wang [12] put forward the concept and method of reduction based on covering rough sets. Bonikowski [10] studied the structures of coverings. Mordeson [13] examined the relationship between the approximations of sets; Chen [14] discussed the covering rough set within the framework of a complete completely distributive lattice. Li [15] put forward a new method of knowledge reduction based on a covering decision system. Xia [16] came up with attribute reduction in covering information systems from information theory. The above researches basically established the theoretical basis and frame of knowledge acquisition and gave effective attribute reduction methods based on covering approximation space, but these researches do not have specific methods for determining the cover.

With the development in information technology, people from all walks of life have accumulated a large amount of data. However, how to get valuable information from these data is one of our key topics. The existing attribute reduction methods based on a covering decision system are very complex and cannot be easily applied to practical problems. Therefore, it is very important to obtain a practical and feasible method of attribute reduction. Over a large number of practical problems, because of the complexity of the objective things and human thought fuzziness, the outside edge of attributes of covering decision system is not clear. Aiming at resolving the reduction problem of quantity characteristics information system, we first put forward the idea of fuzzy attribute; second, we determine the cover through presetting a δ value and give a method of attribute reduction by using the information entropy; finally, we compare our method with [15] through a practical case and prove that our reduction method is more practical than that of [15], and attribute reduction results vary with the standard of covering.

2 Preliminaries

Firstly, we review basic concepts related to covering information system which can be found in [13–17].

Let U be a nonempty, universe of discourse set and \mathcal{B} be a family of subsets of U . If $U = \cup\{A|A \in \mathcal{B}\}$, then \mathcal{B} is called a covering of U , and (U, \mathcal{B}) is called a covering approximation space.

Concept approximation is the key for knowledge acquisition in covering approximate space. The existing researches are generally based on the continuous attribute value and discrete attribute value. Over a very large number of practical problems, condition attributes in covering decision system are uncertain. For attribute “scientific research level,” the attribute values are “high,” “medium,” and “poor.” Obviously, the three attribute values are transitional, and there exist intersections between them. For such a problem, we express “scientific research ability” in terms of fuzzy attribute “scientific research ability high.” A method of determining the covering of fuzzy attribute is stated as follows.

Let U be a finite nonempty set called a universe, A be a fuzzy set on U , $A(x)$ be the membership function of A , $\delta \in [0, 1]$, and $A_{(x, \delta)} = \{y| |A(x) - A(y)| \leq \delta\}$. Then, we say that

$$\mathcal{B}(A, \delta) = \{A_{(x, \delta)}|x \in U\} \tag{1}$$

is the δ -fuzzy covering of U about A [11].

Let $R = \{(x, y)| |A(x) - A(y)| \leq \delta\}$. It is easy to see that R is a fuzzy relation on U . In particular, $\mathcal{B}(A, 0)$ is the partition determined by the membership degree of A and $\mathcal{B}(A, 1) = \{U\}$. For a fuzzy set A , different δ will cause different cover. When we consider multiple fuzzy attributes, there is the following definition:

Let U be a finite nonempty universe, $A^{(1)}, A^{(2)}, \dots, A^{(n)}$ be the fuzzy sets on U and

$$B_x(\{A^{(i)}\}_{i=1}^n, \delta) = \bigcap_{i=1}^n A_{(x, \delta)}^{(i)}, \tag{2}$$

$$\mathcal{B}(\{A^{(i)}\}_{i=1}^n, \delta) = \{B_x(\{A^{(i)}\}_{i=1}^n, \delta)|x \in U\}. \tag{3}$$

Then, we say that $\mathcal{B}(\{A^{(i)}\}_{i=1}^n, \delta)$ is the δ -fuzzy covering of U about $A^{(1)}, A^{(2)}, \dots, A^{(n)}$.

In the following, we define $\mathcal{A} = \{A^{(i)}\}_{i=1}^n$ for simplicity. There may exist many coverings for the universe of discourse U once given a general binary relation and also the comparison problem of the fine relationship among the covers.

Let (U, \mathcal{B}) be a covering approximation space, $x \in U$. Then, we say that $Md_{\mathcal{B}}(x) = \{A|A \in \mathcal{B} \text{ and satisfies the following conditions: (1) } x \in A; (2) \text{ there is no } S \in \mathcal{B} \text{ such that } x \in S, \text{ and } S \subset A, S \neq A\}$ is the minimal description sets of x about (U, \mathcal{B}) . Let $\mathcal{B}_1, \mathcal{B}_2$ be coverings of U , if for any $x \in U$ and $A' \in Md_{\mathcal{B}_1}(x)$

there is a $A'' \in Md_{\mathcal{B}_2}(x)$ such that $A' \subset A''$, then we say \mathcal{B}_1 is finer than \mathcal{B}_2 , and it is denoted by $\mathcal{B}_1 \leq \mathcal{B}_2$ [18].

Property 1 Let U be a finite nonempty universe, A be a fuzzy set on U , $\delta_1, \delta_2 \in [0, 1]$ and $\delta_1 < \delta_2$. Then $\mathcal{B}(A, \delta_1) \leq \mathcal{B}(A, \delta_2)$.

Property 2 Let U be a finite nonempty universe, $A^{(1)}, A^{(2)}, \dots, A^{(n)}$ be the fuzzy sets on U , $\delta_1, \delta_2 \in [0, 1]$ and $\delta_1 < \delta_2$. Then $\mathcal{B}(\mathcal{A}, \delta_1) \leq \mathcal{B}(\mathcal{A}, \delta_2)$.

In the actual problems, by setting δ value, we could control the degree of cover. This operation is easier to understand.

One of the main applications of covering rough set model is attribute reduction. Here, the “reduction” refers to that the covers remain unchanged after we remove a fuzzy attribute.

Let $(U, \mathcal{B}(\mathcal{A}, \delta))$ be a covering information system, for $A^{(j)} \in \mathcal{A}$, if

$$\mathcal{B}(\mathcal{A}, \delta) = \mathcal{B}(\mathcal{A} - \{A^{(j)}\}, \delta), \tag{4}$$

then $A^{(j)}$ is called dispensable in \mathcal{A} otherwise $A^{(j)}$ is called indispensable.

For every $P \subseteq \mathcal{A}$ satisfying $\mathcal{B}(\mathcal{A}, \delta) = \mathcal{B}(P, \delta)$, if every fuzzy set in P is indispensable, that is, for every $A^{(j)} \in P$, $\mathcal{B}(P, \delta) \neq \mathcal{B}(P - \{A^{(j)}\}, \delta)$, then P is called a reduction of \mathcal{A} . The collection of all the indispensable fuzzy sets in \mathcal{A} is called the core of \mathcal{A} , denoted as $Core(\mathcal{A})$.

In the following, we give an information measure for the discernibility of a covering, which is equivalent to Shannon’s entropy [18] if $\mathcal{B}(\mathcal{A}, \delta)$ is a partition.

Definition 1 Define the self-information quantity of the covering class $B_x(\mathcal{A}, \delta)$ by

$$I(B_x(\mathcal{A}, \delta)) = -\log \frac{|B_x(\mathcal{A}, \delta)|}{|U|}, \tag{5}$$

where $|A|$ denotes the number of elements in A .

Definition 2 Let $\mathcal{B}(\mathcal{A}, \delta)$ be the δ -fuzzy covering of U about $A^{(1)}, A^{(2)}, \dots, A^{(n)}$, the information entropy of $\mathcal{B}(\mathcal{A}, \delta)$ is defined as follows:

$$H(\mathcal{B}(\mathcal{A}, \delta)) = -\frac{1}{|U|} \sum_{x \in U} \log_2 \frac{|B_x(\mathcal{A}, \delta)|}{|U|}. \tag{6}$$

For $P, Q \subseteq \mathcal{A}$, the conditional entropy of $\mathcal{B}(P, \delta)$ and $\mathcal{B}(Q, \delta)$ is defined as follows:

$$H(\mathcal{B}(P, \delta) | \mathcal{B}(Q, \delta)) = -\frac{1}{|U|} \sum_{x \in U} \log_2 \frac{|B_x(P, \delta) \cap B_x(Q, \delta)|}{|B_x(Q, \delta)|}. \tag{7}$$

3 Attribute Reduction

Theorem 1 Let $(U, \mathcal{B}(\mathcal{A}, \delta))$ be a covering information system, $\delta \in [0, 1], A^{(j)} \in \mathcal{A}$ is redundant if and only if

$$H(\mathcal{B}(A^{(j)}, \delta) | \mathcal{B}(\mathcal{A} - \{A^{(j)}\}, \delta)) = 0. \tag{8}$$

Proof If $A^{(j)}$ is redundant in \mathcal{A} , then $\mathcal{B}(\mathcal{A}, \delta) = \mathcal{B}(\mathcal{A} - \{A^{(j)}\}, \delta)$, that is, for any $x \in U, B_x(\mathcal{A}, \delta) = B_x(\mathcal{A} - \{A^{(j)}\}, \delta)$. So $H(\mathcal{B}(\mathcal{A}, \delta) | \mathcal{B}(\mathcal{A} - \{A^{(j)}\}, \delta)) = 0$.

Conversely, suppose $H(\mathcal{B}(A^{(j)}, \delta) | \mathcal{B}(\mathcal{A} - \{A^{(j)}\}, \delta)) = 0$. We have

$$\log_2 \frac{|B_x(\mathcal{A} - \{A^{(j)}\}, \delta) \cap B_x(A^{(j)}, \delta)|}{|B_x(\mathcal{A} - \{A^{(j)}\}, \delta)|} \leq 0$$

for any $x \in U$, that is, $H(\mathcal{B}(A^{(j)}, \delta) | \mathcal{B}(\mathcal{A} - \{A^{(j)}\}, \delta)) \geq 0$. If there exists $x_0 \in U$ such that $B_{x_0}(\mathcal{A} - \{A^{(j)}\}, \delta) \cap B_{x_0}(A^{(j)}, \delta) \subset B_{x_0}(\mathcal{A} - \{A^{(j)}\}, \delta)$, then

$$\log_2 \frac{|B_{x_0}(\mathcal{A} - \{A^{(j)}\}, \delta) \cap B_{x_0}(A^{(j)}, \delta)|}{|B_{x_0}(\mathcal{A} - \{A^{(j)}\}, \delta)|} < 0.$$

It follows that $H(\mathcal{B}(A^{(j)}, \delta) | \mathcal{B}(\mathcal{A} - \{A^{(j)}\}, \delta)) > 0$ which is a contradiction. So for all $x \in U, B_x(\mathcal{A} - \{A^{(j)}\}, \delta) \cap B_x(A^{(j)}, \delta) \subset B_x(\mathcal{A} - \{A^{(j)}\}, \delta)$, that is, $\mathcal{B}(\mathcal{A}, \delta) = \mathcal{B}(\mathcal{A} - \{A^{(j)}\}, \delta)$. So $A^{(j)}$ is redundant in \mathcal{A} .

Corollary 1 Let $(U, \mathcal{B}(\mathcal{A}, \delta))$ be a covering information system, $\delta \in [0, 1], A^{(j)} \in \mathcal{A}$ is indispensable if and only if

$$H(\mathcal{B}(A^{(j)}, \delta) | \mathcal{B}(\mathcal{A} - \{A^{(j)}\}, \delta)) > 0. \tag{9}$$

Theorem 2 Let $(U, \mathcal{B}(\mathcal{A}, \delta))$ be a covering information system, $\delta \in [0, 1], P \subseteq \mathcal{A}$ is a reduct of \mathcal{A} if and only if (1) $H(\mathcal{B}(\mathcal{A}, \delta)) = H(\mathcal{B}(P, \delta))$; (2) For any $A^{(j)} \in P, H(\mathcal{B}(A^{(j)}, \delta) | \mathcal{B}(P - \{A^{(j)}\}, \delta)) > 0$.

Proof If $P \subseteq \mathcal{A}$ is a reduct of \mathcal{A} . Let $Q = \mathcal{A} - P$, then $H(\mathcal{B}(Q, \delta) | \mathcal{B}(P, \delta)) = 0$, that is, $\log(|B_x(Q, \delta) \cap B_x(P, \delta)| / |B_x(P, \delta)|) = 0$. Hence, $B_x(P, \delta) = B_x(\mathcal{A}, \delta)$ for any $x \in U$. So $H(\mathcal{B}(\mathcal{A}, \delta)) = H(\mathcal{B}(P, \delta))$. Because $P \subseteq \mathcal{A}$ is a reduct of \mathcal{A} , for any $A^{(j)} \in P, A^{(j)}$ is indispensable, that is, $H(\mathcal{B}(A^{(j)}, \delta) | \mathcal{B}(P - \{A^{(j)}\}, \delta)) > 0$.

Conversely, suppose $H(\mathcal{B}(\mathcal{A}, \delta)) = H(\mathcal{B}(P, \delta))$, and for any $A^{(j)} \in P, H(\mathcal{B}(A^{(j)}, \delta) | \mathcal{B}(P - \{A^{(j)}\}, \delta)) > 0$. Then for any $A^{(j)} \in P, A^{(j)}$ is indispensable, and $B_x(P, \delta) = B_x(\mathcal{A}, \delta)$. So $H(\mathcal{B}(\mathcal{A} - P, \delta) | \mathcal{B}(P, \delta)) = 0$. Hence, $P \subseteq \mathcal{A}$ is a reduct of \mathcal{A} .

4 Example Analysis

Faced with increasingly fierce competition in the market today, the enterprises feel that bringing into full play the human resources is an important means of business development. To carry out effective employee training and development management is the premise condition for giving full play to human resources. Before the commencement of training activities, there is an analysis of employees' demand. According to the different levels of staff, the enterprise will carry out appropriate training. But in many cases, employees may belong to different classes, so we need to use cover to solve the problem.

There are nine workers $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\}$ in a department. For group training, the company sends five experts E_i ($i = 1, 2, 3, 4, 5$) to review the workers through five aspects: *Business level*, *Communication ability*, *Work attitude*, *Innovative ability*, and *Learning ability*. Every expert will choose four workers who are more accordance with the condition. We have five specialists to evaluate the attribute values for these workers in Table 1. Where the fuzzy attribute "Business level higher," "Communication ability better," "Work steadfast," "Innovative ability," and "Quick learner" are denoted by $A^{(1)}$, $A^{(2)}$, $A^{(3)}$, $A^{(4)}$, and $A^{(5)}$, and $\mathcal{A} = \{A^{(1)}, A^{(2)}, A^{(3)}, A^{(4)}, A^{(5)}\}$.

To obey the principle of fair justice, experts choose four workers who are more accordance with the condition. Through Table 1, we get the membership degrees of these workers according to these fuzzy sets (Table 2). For example, there are five experts and four of them consider x_1 fit for the business level higher ($A^{(1)}$). Then we think the membership degrees of x_1 fit for the business level higher ($A^{(1)}$) is 4/5.

If we want to divide the high degree of similar staff into a same class, we make $\delta = 0.2$. According to the method given by section two, by using the data in Table 2, we get the δ -fuzzy coverings of the universe of discourse U based on the five fuzzy attributes (Tables 3 and 4). Among them, \hat{x}_i represent that x_i is the core. And the set is get by formula $A_{(x,\delta)} = \{y \mid |A(x) - A(y)| \leq \delta\}$. Like $A_{(x_1,\delta)}^{(1)} = A_{(x_3,\delta)}^{(1)} = \{\hat{x}_1, x_2, \hat{x}_3, x_4, x_8\}$.

The meaning of $\delta = 0.2$ is to divide the workers whose degree of similarity no less than 0.8 into a same class. It helps enterprise develop a more targeted training.

Table 1 The attribute values of these workers

Experts	Fuzzy Attributes				
	$A^{(1)}$	$A^{(2)}$	$A^{(3)}$	$A^{(4)}$	$A^{(5)}$
E_1	x_2, x_3, x_4, x_8	x_1, x_2, x_3, x_4	x_1, x_2, x_3, x_4	x_1, x_3, x_4, x_7	x_1, x_3, x_4, x_7
E_2	x_1, x_2, x_3, x_4	x_1, x_3, x_5, x_8	x_1, x_2, x_3, x_7	x_2, x_3, x_4, x_9	x_2, x_3, x_4, x_7
E_3	x_1, x_2, x_4, x_8	x_1, x_3, x_4, x_5	x_1, x_3, x_4, x_7	x_1, x_3, x_4, x_6	x_2, x_3, x_4, x_6
E_4	x_1, x_2, x_3, x_8	x_2, x_3, x_4, x_8	x_1, x_2, x_4, x_7	x_1, x_4, x_5, x_9	x_1, x_3, x_7, x_9
E_5	x_1, x_2, x_3, x_6	x_1, x_2, x_3, x_5	x_1, x_2, x_4, x_9	x_2, x_3, x_4, x_5	x_2, x_3, x_4, x_9

Table 2 The membership degrees

Fuzzy attributes	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
$A^{(1)}$	4/5	1	4/5	3/5	0	1/5	0	3/5	0
$A^{(2)}$	4/5	3/5	1	3/5	3/5	0	0	2/5	0
$A^{(3)}$	1	4/5	3/5	4/5	0	0	3/5	0	1/5
$A^{(4)}$	3/5	2/5	4/5	1	2/5	1/5	1/5	0	2/5
$A^{(5)}$	2/5	3/5	1	4/5	0	1/5	3/5	0	2/5

Table 3 The fuzzy coverings for $\delta = 0.2$

Fuzzy concepts	Fuzzy coverings
$A^{(1)}$	$\{\hat{x}_1, x_2, \hat{x}_3, x_4, x_8\}, \{x_1, \hat{x}_2, x_3\}, \{\hat{x}_5, \hat{x}_6, \hat{x}_7, \hat{x}_9\}, \{x_1, x_3, x_4, \hat{x}_8\}$
$A^{(2)}$	$\{\hat{x}_1, x_2, x_3, x_4, x_5\}, \{x_1, \hat{x}_2, \hat{x}_4, \hat{x}_5, x_8\}, \{x_1, \hat{x}_3\}, \{\hat{x}_6, \hat{x}_7, \hat{x}_9\}, \{x_2, x_4, x_5, \hat{x}_8\},$ $\{\hat{x}_1, x_2, x_4\}, \{x_1, \hat{x}_2, x_3, \hat{x}_4, x_7\}, \{x_2, \hat{x}_3, x_4, \hat{x}_7\}, \{\hat{x}_5, \hat{x}_6, \hat{x}_8, \hat{x}_9\}$
$A^{(4)}$	$\{x_1, \hat{x}_3, x_4\}, \{x_3, \hat{x}_4\}, \{\hat{x}_1, x_2, x_3, x_5, x_9\}, \{x_6, x_7, \hat{x}_8\}, \{x_2, x_5, \hat{x}_6, \hat{x}_7, x_8, x_9\},$ $\{x_1, \hat{x}_2, \hat{x}_5, x_6, x_7, \hat{x}_9\}$
$A^{(5)}$	$\{x_3, \hat{x}_4\}, \{x_2, x_3, \hat{x}_4, x_7\}, \{\hat{x}_1, x_2, x_6, x_7, \hat{x}_9\}, \{x_1, \hat{x}_2, x_4, \hat{x}_7, x_9\}, \{x_1, x_5, \hat{x}_6, x_8, x_9\},$ $\{\hat{x}_5, x_6, \hat{x}_8\}, \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9\},$
A	$\{\hat{x}_1, \hat{x}_2\}, \{\hat{x}_3\}, \{\hat{x}_4\}, \{\hat{x}_5\}, \{\hat{x}_6, \hat{x}_9\}, \{\hat{x}_7\}, \{\hat{x}_8\}$

For comparison, we can take $\delta = 0.4$ to get a different covers. Through Tables 3 and 4, we know that when $\delta_1 < \delta_2, \mathcal{B}(A^{(i)}, \delta_1) \leq \mathcal{B}(A^{(i)}, \delta_2), i = 1, 2, 3, 4, 5,$ and $\mathcal{B}(\mathcal{A}, \delta_1) \leq \mathcal{B}(\mathcal{A}, \delta_2).$ In different situation, the company could combine staffs in different classes.

According to the method given in Sect. 3 and the data in Tables 3 and 4, we get the reduction results as shown in Table 5: When we take $\delta = 0.2,$ the core attributes are $A^{(2)}, A^{(3)}, A^{(4)}.$ It means that we could only consider “Communication ability better,” “Work steadfast,” and “Innovative ability” in the expert evaluation. The other two fuzzy attributes cannot be considered. This makes the enterprise to reduce the work and enhance the working efficiency. It has some practical significance.

Through Table 5, we know that attribute reduction results vary with the standard of covering. Comparing the two kinds of reduction method, when δ is smaller, that is, the demand of similarity level is high, reduction degree of two reduction methods is same; when δ is bigger, that is, the demand of similarity level is low, the reduction results of [15] are more concise.

Essentially, the Δ from [15] is not monotonic, and $\mathcal{B}(\mathcal{A}, \delta)$ given by our paper is monotonic. From the practical point of view, the reduction method in [15] was cumbersome. The induced covering was too fine to intuitively describe the problems. In practical problems, getting the core attribute is our ultimate goal, so there is no need making coverings too fine. In this paper, using $\mathcal{B}(\mathcal{A}, \delta) = \{\mathcal{B}_x(\mathcal{A}, \delta) | x \in U\}$ to describe issues will be more intuitive. As long as according to actual situation needs to choose the appropriate level of similarity degree $\delta,$ we can find the core attributes. In different situation, company could combine staffs in different classes.

Table 4 The fuzzy coverings for $\delta = 0.4$

	Fuzzy coverings
$A^{(1)}$	$\{\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{x}_4, \hat{x}_8\}, \{x_1, x_2, x_3, \hat{x}_4, x_6, \hat{x}_8\}, \{\hat{x}_5, x_6, \hat{x}_7, \hat{x}_9\}, \{x_4, x_5, \hat{x}_6, x_7, x_8, x_9\}$
$A^{(2)}$	$\{\hat{x}_1, \hat{x}_2, x_3, \hat{x}_4, \hat{x}_5, x_8\}, \{x_1, x_2, \hat{x}_3, x_4, x_5\}, \{\hat{x}_6, \hat{x}_7, x_8, \hat{x}_9\}, \{x_1, x_2, x_4, x_5, x_6, x_7, \hat{x}_8, x_9\}$
$A^{(3)}$	$\{\hat{x}_1, \hat{x}_2, x_3, \hat{x}_4, x_7\}, \{x_1, x_2, \hat{x}_3, x_4, \hat{x}_7, x_9\}, \{\hat{x}_5, \hat{x}_6, \hat{x}_8, x_9\}, \{x_3, x_5, x_6, x_7, x_8, \hat{x}_9\}$
$A^{(4)}$	$\{x_1, x_3, \hat{x}_4\}, \{x_2, x_5, x_6, x_7, \hat{x}_8, x_9\}, \{x_1, x_2, x_5, \hat{x}_6, \hat{x}_7, x_8, x_9\}, \{\hat{x}_1, x_2, x_3, x_4, x_5, x_6, x_7, x_9\},$ $\{x_1, \hat{x}_2, x_3, \hat{x}_5, x_6, x_7, x_8, \hat{x}_9\}, \{x_1, x_2, x_4, x_5, x_6, x_7, x_8, \hat{x}_9\}, \{x_1, x_2, \hat{x}_3, x_4, x_5, x_6, x_7, x_9\},$
$A^{(5)}$	$\{x_1, \hat{x}_5, x_6, \hat{x}_8, x_9\}, \{x_2, \hat{x}_3, x_4, x_7\}, \{\hat{x}_1, x_2, x_4, x_5, x_6, x_7, x_8, \hat{x}_9\}, \{x_1, \hat{x}_2, x_3, x_4, x_6, \hat{x}_7, x_9\},$ $\{x_1, \hat{x}_2, x_3, x_4, x_6, \hat{x}_7, x_9\}, \{x_1, x_2, x_3, \hat{x}_4, x_7, x_9\}, \{x_1, x_2, x_5, \hat{x}_6, x_7, x_8, x_9\}$
\mathcal{A}	$\{\hat{x}_1, x_2, x_4\}, \{x_1, \hat{x}_2, x_3\}, \{x_6, \hat{x}_8\}, \{x_2, \hat{x}_3, x_4\}, \{x_1, x_3, \hat{x}_4\}, \{x_1, x_3, x_4\}, \{\hat{x}_5\}, \{\hat{x}_6, x_8, x_9\}, \{\hat{x}_7, x_9\}, \{\hat{x}_7, x_9\}$

Table 5 Reduction results

Methods	Reduction results ($\delta = 0.2$)	Reduction results ($\delta = 0.4$)	Reduction results ($\delta = 0.6$)
This paper	$\{A^{(2)}, A^{(3)}, A^{(4)}\}$	$\{A^{(1)}, A^{(2)}, A^{(3)}, A^{(4)}, A^{(5)}\}$	$\{A^{(1)}, A^{(2)}, A^{(3)}, A^{(4)}, A^{(5)}\}$
Literature[15]	$\{A^{(2)}, A^{(3)}, A^{(4)}\}$	$\{A^{(1)}, A^{(2)}, A^{(4)}, A^{(5)}\}$	$\{A^{(1)}, A^{(2)}, A^{(3)}, A^{(4)}\}$

5 Conclusion

Based on the concept of covering and the foundation of rough set model put forward by Zakowski, in this paper, we put forward the idea of fuzzy attribute. Focusing on practical problems, we discuss an effective attribute reduction method which is more intuitive than other methods. We could determine the cover through setting a δ value and further give a method of attribute reduction by using the information entropy. Finally, we compared our method with [15] through a practical case, and we proved that our reduction method is more practical. Meanwhile, we also analyze the impact of precision change on coverings ascertained by fuzzy sets and verify that attribute reduction results vary with the standard of covering. These results provide solid foundation for the application of covering-based rough set models to fuzzy attributes.

Acknowledgments This work is supported by the National Natural Science Foundation of China (71071049) and the Natural Science Foundation of Hebei Province (F2011208056).

References

1. Pawlak Z (1998) Rough sets theory and its applications to data analysis. *Cybern Syst* 29:661–668
2. Ziarko W (1993) Variable precision rough set model. *J Comput Syst Sci* 46:39–59
3. Dubois D, Prade H (1992) Upper and lower images of a fuzzy set induced by a fuzzy relation. *Inf Sci* 64(3):203–232
4. Dubois D, Prade H (1990) Rough fuzzy sets and fuzzy rough sets. *J Gen Syst* 17:191–208
5. Morsi N, Yakout M (1998) Axiomatics for fuzzy rough sets. *Fuzzy Sets Syst* 100:327–342
6. Weizhi W, Jusheng M, Wenxiu Z (2003) Generalized fuzzy rough sets. *Inf Sci* 151:263–282
7. Jusheng M, Yee L, Huiyin Z, Tao F (2008) Generalized fuzzy rough sets determined by a triangular norm. *Inf Sci* 178:3203–3213
8. Zadeh LA (1965) Fuzzy sets. *Inf Control* 8:338–353
9. Zakowski W (1983) Approximations in the space. *Demonstrate Math* 16:761–769
10. Bonikowski Z (1998) Extensions and intention in the rough set theory. *Inf Sci* 107:149–167
11. Jun Z, Qingling Z, Wenshi C (2004) The fuzzy set subsystem based on the equivalent class. *J Northeast Univ* 25:731–733
12. William Z, Feiyue W (2003) Reduction and maximization of covering generalized rough sets. *Inf Sci* 152:217–230
13. Mordeson JN (2001) Rough set theory applied to ideal theory. *Fuzzy Sets Syst* 121:315–324

14. Degang C, Changzhong W, Qinghua H (2007) A new approach to attribute reduction of consistent and inconsistent covering decision systems with covering rough sets. *Inf Sci* 177:3500–3518
15. Fei L (2009) Approach to knowledge reduction of covering decision systems based on information theory. *Inf Sci* 179:1694–1704
16. Xiuyun X, Keyun Q (2011) Some notes on attribute reduction based on inconsistent covering decision system. *Comput Eng Appl* 47:97–101
17. Jun H, Guoyin W (2008) Hierarchical model of covering granular space. *J Nanjing Univ* 44:551–558
18. Qinghua H (2006) Fuzzy probabilistic approximation spaces and their information measures. *IEEE Trans Fuzzy Syst* 14:191–201

Sliding Mode-Like Fuzzy Logic Control with Boundary Layer Self-Tuning for Discrete Nonlinear Systems

Xiaoyu Zhang and Fang Guo

Abstract This paper presented a new sliding mode-like fuzzy logic control design for discrete nonlinear systems. Firstly, the boundary layer is self-tuned online, and then, the chattering free is obtained. Consequently, the fuzzy logic control (FLC) is designed to approximate the sliding mode control (SMC) with boundary layer self-tuning. Finally, the performance of the robustness, chattering free, and adaption is verified by the simulation results.

Keywords Sliding mode control · Chattering free · Fuzzy logic system · Adaptive

1 Introduction

Sliding mode control (SMC) is investigated by many researchers especially in recent years. Its main theory is discussed in [1–3]. As one of the main approaches to control design for nonlinear systems, SMC has many advantages such as robustness, invariance to system uncertainties, and resistance to the external disturbance. However, the obvious shortcoming of SMC is the chattering phenomenon which impedes the application. Fortunately, many papers about chattering free problem are proposed such as [4–6].

Fuzzy logic control (FLC) is widely applied especially to the plant in which precise mathematics model cannot be acquired easily. Some researchers apply fuzzy logic systems to sliding mode control to improve the performance of SMC. Hence, there are two combined methods: fuzzy sliding mode control (FSMC, [7–9]) and sliding mode fuzzy logic control (SMFC, [10–14]). The former is fuzzy

X. Zhang (✉) · F. Guo

Department of Electronics and Information Engineering, North China Institute of Science and Technology, Beijing, China

e-mail: ysuzxy@aliyun.com; ysuzxy@163.com

adaptive sliding mode control algorithm in which unknown system dynamics are identified by FLS to form the equivalent control of SMC. The latter is FLC design based on sliding mode or sliding mode control.

Lhee et al. put forward sliding-like fuzzy logic control (SMLFC) [10, 11]. In their method, fuzzy controller is equivalent to the sliding mode controller pre-designed and the boundary layer thickness can be self-tuned online after introducing appropriate adaptive laws, wherefore the controller has no chattering. However, in papers [10, 11], the methods are only applicable to systems with second-order dynamics. Furthermore, as the dead zone parameter converges to zero, signal output of fuzzy controller also has a little chattering. For discrete nonlinear systems, there rarely are reported research works of the SMLFC with boundary layer self-tuning by so far.

Sliding mode-like fuzzy logic control (SMLFC) for discrete nonlinear systems is mainly considered in this paper. Based on the reported works [10, 11, 14, 15], controller design is enhanced to be applicable to discrete nonlinear systems with nonlinear control gains. Dynamic fuzzy logic systems (DFLS) are also utilized for the more smoothing performance, in which parameters can be regulated online to acquire better performance. As the same idea in paper [10, 11, 14, 15], the FLC controller, which is equivalent to the pre-designed SMC controller, is also obtained finally. Chattering phenomenon will be eliminated completely. This paper is organized as follows. In Sect. 2, problem formulation and general SMC are presented. SMC with boundary layer thickness self-tuning is approached in Sect. 3, and the reachability of the sliding mode will be verified by Lyapunov stability theory. Then, the FLC controller which is equivalent to the pre-designed SMC is reported in Sect. 4. Simulation results of numerical examples are proposed in Sect. 5 to validate the controller design. Conclusions are summarized in Sect. 6.

2 Problem Formula and SMC

Consider a class of discrete nonlinear systems,

$$x_i(k+1) = x_{i+1}(k) \quad (i = 1, 2, \dots, n-1), \quad x_n(k+1) = f(x, k) + g(x, k)u(k) \quad (1)$$

where $x = [x_1(k) \dots x_n(k)]^T \in R^n$ is the state vector, $f(x, k)$ is the nonlinear dynamic, $u(k)$ is the control input, $g(x, k)$ is the nonlinear gain of $u(k)$.

It is assumed that the desired trajectory is y_d and $x_d = [x_{d1}(k) \dots x_{dn}(k)]^T$ is the corresponding auxiliary state vector. Suppose that $y_d = x_{d1}(k)$ and $x_{di}(k)$ ($i = 1, 2, \dots, n$) is bounded for all time interval k ($k \in [0, +\infty)$) which satisfies the stable tracked model

$$\begin{aligned} x_{di}(k+1) &= x_{d(i+1)}(k) \quad (i = 1, 2, \dots, n-1), \\ x_{dn}(k+1) &= - \sum_{i=0}^{n-1} a_i x_{d(i+1)}(k) + r(k) \end{aligned} \quad (2)$$

The problem is to construct sliding mode control law $u(r(k), x_d(k), x(k), s(k), k)$ such that the error vector

$$e = [e_1(k) \dots e_n(k)]^T, \quad e_i(k) = x_i(k) - x_{di}(k), \quad i = 1, 2, \dots, n \quad (3)$$

converges to the tolerable range, where $s(k)$ is the sliding mode surface

$$s(k) = c^T e, \quad c = [c_1 \dots c_{n-1}]^T \quad (4)$$

and the vector c make a Hurwitz polynomial.

Suppose that the nonlinear dynamic function f can be estimated as $\hat{f}(x, k)$, and the nonlinear control gain g can be estimated as \hat{g} by designers where \hat{g} is a constant scalar. Then, the nonlinear uncertainty Δf and Δg are defined as

$$\Delta f = f(x, k) - \hat{f}(x, k), \quad \Delta g = g(x, k) - \hat{g}. \quad (5)$$

Without loss of generality, suppose that $g(x, k) > 0$, $\hat{g} > 0$, and $\Delta \bar{g}$ is the upper boundary of Δg . Then, $\hat{g}^{-1} \Delta g > -1$ holds.

For convenience define that

$$\delta = |\Delta f| + |\Delta g \times u_{eq}|, \quad \gamma = 1 + \hat{g}^{-1} \Delta g. \quad (6)$$

And then $\delta > 0$, $0 < \gamma \leq \bar{\gamma}$ with $\bar{\gamma} = 1 + \hat{g}^{-1} \Delta g$. Here, u_{eq} is the equivalent control of the sliding mode control $u(k)$.

According to the basic sliding mode theory in [1] and parts of the content in [10, 11], if the SMC control input u is constructed as

$$\begin{aligned} u(k) &= u_{eq} + u_v \\ u_{eq} &= \hat{g}^{-1} \left[cx_d(k+1) - \sum_{i=1}^{n-1} c_i x_{i+1}(k) - \hat{f}(x, k) + s(k) \right] \\ u_v &= -\hat{g}^{-1} k_1 \text{sat} \left(\frac{s(k)}{\varphi} \right) \end{aligned} \quad (7)$$

where

$$\text{sat} \left(\frac{s}{\varphi} \right) = \begin{cases} s/\varphi, & \text{if } 0 < |s| < \varphi \\ \text{sgn}(s), & \text{if } |s| \geq \varphi \\ 0, & \text{if } s = 0 \end{cases} \quad (8)$$

and

$$\varphi = \delta + \eta + \varepsilon \quad (9)$$

is the thickness of the boundary layer [11–13] with $\eta > 0$ and $\varepsilon > 0$ are positive scalars that can be optionally small. k_1 is given by

$$k_1 = \gamma^{-1} (\delta + \eta) \quad (10)$$

Then, the reaching condition

$$\Delta[s^2(k)] \leq -\eta|\Delta s(k)| \tag{11}$$

can be satisfied and the system tracking error vector e is bounded.

Theorem 1 For the error system (1)–(3) and its sliding mode (4), the reaching condition (10) can be satisfied and the switching area $\mathbf{B} = \{e(k)|s(k)| \leq \varphi\}$ is also reached and stable, if the control is designed as (7), (8). Furthermore, $s(k)$ can be driven to $\mathbf{S} = \{e(k)|s(k) = 0\}$ if $\Delta f = \Delta g = 0$.

Proof The Lyapunov function $V(k) = s^2(k)$ is selected and then its difference

$$\Delta V(k) = V(k + 1) - V(k) = s^2(k + 1) - s^2(k) = \Delta s(k)[\Delta s(k) + 2s(k)],$$

is derived, whereas (11) equals

$$\Delta s(k)\{\Delta s(k) + 2s(k) + \eta \text{sgn}[\Delta s(k)]\} < 0. \tag{12}$$

According to (1)–(6),

$$\Delta s(k) = \Delta f + \Delta g u_{\text{eq}} - (\delta + \eta) \text{sat}\left(\frac{s(k)}{\varphi}\right) \tag{13}$$

is obtained certainly.

If $s(k) > \varphi$, let $s(k) = \varphi + \xi_1(k)$ where $\xi_1(k) > 0$, then (12) becomes

$$\underbrace{[\Delta f + \Delta g u_{\text{eq}} - \delta - \eta]}_{\text{negative}} [\Delta f + \Delta g u_{\text{eq}} - \delta - \eta + 2\varphi + 2\xi_1(k) - \eta] < 0.$$

Using equation (6) and (9), the above becomes

$$\underbrace{[\Delta f + \Delta g u_{\text{eq}} - \delta - \eta]}_{\text{negative}} \underbrace{[\Delta f + \Delta g u_{\text{eq}} + \delta + 2\varepsilon + 2\xi_1(k)]}_{\text{positive}} < 0,$$

which implies that inequality (12) is satisfied.

If $s(k) < -\varphi$, let $s(k) = -\varphi - \xi_1(k)$ where $\xi_1(k) < 0$, then (12) becomes

$$[\Delta f + \Delta g u_{\text{eq}} + \delta + \eta] [\Delta f + \Delta g u_{\text{eq}} + \delta + \eta - 2\varphi - 2\xi_1(k) + \eta] < 0.$$

Using equation (6) and (9), the above becomes

$$\underbrace{[\Delta f + \Delta g u_{\text{eq}} + \delta + \eta]}_{\text{positive}} \underbrace{[\Delta f + \Delta g u_{\text{eq}} - \delta - 2\varepsilon - 2\xi_1(k)]}_{\text{negative}} < 0$$

which implies that inequality (12) is again satisfied. Thus, inequality (11) holds and switching area \mathbf{B} is reachable and stable according to Lyapunov stability theory.

If $\Delta f = \Delta g = 0$, from Eq. (13), one has $s(k + 1) = \left[1 - \frac{(\delta + \eta)}{\varphi}\right]s(k)$, and it is obvious from (9) that $-1 < 1 - \frac{(\delta + \eta)}{\varphi} < 1$ which implies that $s(k)$ is a stable first-order filter that asymptotically approaches $\mathbf{S} = \{e(k)|s(k) = 0\}$. \square

However, parameters γ and δ are all unknown, so the control (7) cannot be implemented because k_1 and φ cannot be determined.

3 SMC with Boundary Layer Self-Tuning

Supposing that k_1 is constant scalar which satisfies the Eq. (9), and if in the Eq. (7) k_1 is replaced by $k_1 - \Delta\varphi(k) - |s(k)|$, then (10) changes to

$$\Delta[s^2(k)] \leq -[\eta - \Delta\varphi(k) - |s(k)|]|\Delta s(k)| \tag{14}$$

Theorem 2 For the error system (1–3) and its sliding mode (4), the reaching condition (14) can be satisfied and the switching area $\mathbf{B}' = \{e(k)|s(k) \leq \varphi(k)\}$ is also reached and stable, if the control is designed as (7, 8). Furthermore, the sliding mode $s(k)$ can be driven to $S' = \{e(k)|s(k) = 0\}$ if $\Delta f = \Delta g = 0$.

Proof According to the proof of theorem 1, the Lyapunov function $V(k) = s^2(k)$ is also used here. And φ is replaced with $\varphi(k)$, η is replaced with $\eta - \Delta\varphi(k) - |s(k)|$, then (11) equals

$$\Delta s(k)\{\Delta s(k) + 2s(k) + [\eta - \Delta\varphi(k) - |s(k)|]\text{sgn}[\Delta s(k)]\} < 0.$$

If $s(k) > \varphi$, let $s(k) = \varphi + \xi_1(k)$ where $\xi_1(k) > 0$, one has

$$\underbrace{[\Delta f + \Delta g u_{eq} - \gamma k_1]}_{\text{negative}} [\Delta f + \Delta g u_{eq} - \gamma k_1 + 2\varphi + 2\xi_1(k) - \eta + \Delta\varphi(k) + |s(k)|] < 0.$$

Obviously, η is replaced with $\eta - \Delta\varphi(k) - |s(k)|$ in (9), (10), the above becomes

$$\underbrace{[\Delta f + \Delta g u_{eq} - \gamma k_1]}_{\text{negative}} \underbrace{[\Delta f + \Delta g u_{eq} + \delta + 2\varepsilon + 2\xi_1(k)]}_{\text{positive}} < 0$$

which implies that inequality (12) is satisfied.

If $s(k) < -\varphi$, let $s(k) = -\varphi - \xi_1(k)$ where $\xi_1(k) < 0$, then (14) becomes

$$\underbrace{[\Delta f + \Delta g u_{eq} + \gamma k_1]}_{\text{positive}} [\Delta f + \Delta g u_{eq} + \gamma k_1 - 2\varphi - 2\xi_1(k) + \eta - \Delta\varphi(k) - |s(k)|] < 0.$$

Using (6), (9), and η is replaced with $\eta - \Delta\varphi(k) - |s(k)|$, the above becomes

$$\underbrace{[\Delta f + \Delta g u_{eq} + \gamma k_1]}_{\text{positive}} \underbrace{[\Delta f + \Delta g u_{eq} - \delta - 2\varepsilon - 2\xi_1(k)]}_{\text{negative}} < 0$$

which implies that inequality (12) is satisfied.

And then, $s(k)$ converges to $\mathbf{B}' = \{e(k) \mid |s(k)| \leq \varphi(k)\}$, where $\varphi(k) = \delta + \eta - \Delta\varphi(k) - |s(k)| + \varepsilon$. Namely

$$|s(k)| \leq \delta + \eta - \Delta\varphi(k) - |s(k)| + \varepsilon \Leftrightarrow 0.5|s(k)| \leq \delta + \eta - \Delta\varphi(k) + \varepsilon,$$

and furthermore

$$\varphi(k) = \delta + \eta - \Delta\varphi(k) + \varepsilon \Leftrightarrow \varphi(k + 1) = \delta + \eta + \varepsilon,$$

thus $|s(k)| \leq \varphi(k)$ is reachable. \square

Now, the reaching condition (14) is modified as

$$\Delta[s^2(k)] \leq -[\eta - \lambda_1 \Delta\varphi(k) - \lambda_2 |s(k)|] |\Delta s(k)| \tag{15}$$

where $\lambda_1, \lambda_2 > 1$ are optional positive scalars. Then, the following theorem is referred.

Theorem 3 For the error system (1)–(3) and its sliding mode (4), the reaching condition (15) can be satisfied and the switching area $\mathbf{B}' = \{e(k) \mid |s(k)| \leq \varphi(k)\}$ is also reached and stable, if the control is designed as (7) and (8). Furthermore, the sliding mode $s(k)$ can be driven to $S' = \{e(k) \mid s(k) = 0\}$ if $\Delta f = \Delta g = 0$.

Proof According to the proof of theorem 2, φ is replaced with $\varphi(k)$ and η is replaced with $\eta - \lambda_1 \Delta\varphi(k) - \lambda_2 |s(k)|$, it is referred that $\mathbf{B}' = \{e(k) \mid |s(k)| \leq \varphi(k)\}$ is reachable and $\varphi(k) = \delta + \eta - \lambda_1 \Delta\varphi(k) - \lambda_2 |s(k)| + \varepsilon$. Namely,

$$|s(k)| \leq \delta + \eta - \lambda_1 \Delta\varphi(k) - \lambda_2 |s(k)| + \varepsilon \Leftrightarrow (1 + \lambda_2)|s(k)| \leq \delta + \eta - \lambda_1 \Delta\varphi(k) + \varepsilon$$

holds and then $\varphi(k) = \delta + \eta - \lambda_1 \Delta\varphi(k) + \varepsilon$, that is,

$$\varphi(k + 1) - \frac{\lambda_1 - 1}{\lambda_1} \varphi(k) = \frac{1}{\lambda_1} (\delta + \eta + \varepsilon). \tag{16}$$

Obviously, $|s(k)| \leq \varphi(k)$ where $\varphi(k) = \frac{1}{\lambda_1} (\delta + \eta + \varepsilon)$. \square

However, k_1 or $k_1 - \Delta\varphi(k)$ is unknown in practice. Define \hat{k}_1 is the estimating value of $k_1 - \Delta\varphi(k)$ and $\tilde{k}_1 = \hat{k}_1 - [k_1 - \Delta\varphi(k)]$ is the error of the estimation, then the condition (11) should be satisfied. Choose the following adaptive law

$$\Delta \hat{k}_1(k) = \begin{cases} \alpha(1 + \beta)^{-1} [\beta \hat{k}_1(k) - \alpha\varphi(k) + s(k)], & |s| > \varphi \\ -\tau |s|, & |s| \leq \varphi \end{cases} \tag{17}$$

$$\Delta\varphi(k) = \begin{cases} -\alpha\varphi(k) + \beta\hat{k}_1(k) + s(k), & |s| > \varphi \\ -2\hat{k}_1(k), & |s| \leq \varphi \end{cases} \quad (18)$$

where $\alpha > 0, \beta > 0, \tau > 1$ are all constant scalars, that is, the filtering parameters of the adaptive law. Here, we assume that $\hat{k}(0) > 0, \varphi(0) > 0$, and then, based on the SMC (7) and the adaptive law (12–13), the following theorem can be approached.

Theorem 4 *For nonlinear system (1) with the control law (7) and the adaptive law (17–18), the sliding mode $s(k)$ reaches to $\mathbf{B}' = \{e(k) \mid |s(k)| \leq \varphi(k)\}$ and the closed-loop system is asymptotically stable if the control gain k_1 in (7) is replaced with \hat{k}_1 by being adaptively tuned online.*

Proof Choose a Lyapunov function as

$$V(k) = s^2(k) + \tilde{k}^2 \quad (19)$$

Then, the derivate of the Lyapunov function V with respect to the time k along the trajectory of the closed-loop system can be obtained as following,

$$\Delta V(k) = s^2(k + 1) - s^2(k) + \tilde{k}_1^2(k + 1) - \tilde{k}_1^2(k),$$

and the inequality (15) consequently becomes

$$\Delta s(k) \{ 2\Delta s(k) + 2s(k) + [\eta - \lambda_1 \Delta\varphi(k) - \lambda_2 |s(k)|] \text{sgn}[\Delta s(k)] \} + \Delta \tilde{k}_1(k) [\Delta \tilde{k}_1(k) + 2\tilde{k}_1(k)] < 0. \quad (20)$$

According to the Eq. (13), one has

$$\Delta s(k) = \Delta f + \Delta g u_{eq} - \gamma k_1 \text{sat}\left(\frac{s(k)}{\varphi}\right) + \gamma(k_1 - \hat{k}_1) \text{sat}\left(\frac{s(k)}{\varphi}\right), \quad (21)$$

and from (17) and (18) one has $\Delta \tilde{k}_1(k) = \Delta \hat{k}_1(k) + \Delta^2 \varphi(k) = \Delta s(k)$. Substitute the above into (20),

$$\Delta s(k) \{ 2\Delta s(k) + 2s(k) + [\eta - \lambda_1 \Delta\varphi(k) - \lambda_2 |s(k)|] \text{sgn}[\Delta s(k)] + 2\tilde{k}_1(k) \} < 0, \quad (22)$$

and let $\lambda_1 = 4, \lambda_2 = 8$ for convenience, then

$$\Delta s(k) \{ \Delta s(k) + s(k) + [0.5\eta - 2\Delta\varphi(k) - 4|s(k)|] \text{sgn}[\Delta s(k)] + \tilde{k}_1(k) \} < 0, \quad (23)$$

and by (16),

$$\varphi(k) = \frac{1}{4}(\delta + \eta + \varepsilon). \quad (24)$$

In the following, $s(k) > \varphi(k)$ and $s(k) < -\varphi(k)$ are considered correspondingly.

(1) when $s(k) > \varphi(k)$.

(a) If $\Delta s(k) > 0$, by (10), (18), and (24), the inequality (23) becomes

$$\Delta s(k) \left\{ \underbrace{\Delta f + \Delta g u_{eq} - \gamma k_1}_{\text{negative}} + (1 - \gamma - \beta) \hat{k}_1 + 0.5\eta + (\gamma - 1)(\delta + \eta) + \frac{\alpha}{4}(\delta + \eta + \varepsilon) - 4|s(k)| \right\} < 0$$

Because $\beta > 1 + \bar{\gamma}$ and $\alpha < -4(1 + \bar{\gamma})$,

$$\Delta s(k) \left\{ \underbrace{\Delta f + \Delta g u_{eq} - \gamma k_1}_{\text{negative}} + \underbrace{(1 - \gamma - \beta) \hat{k}_1}_{\text{negative}} - \underbrace{4|s(k)|}_{\text{negative}} + \underbrace{(\gamma - 1)\delta + (\gamma - 0.5)\eta + \alpha(\delta + \eta + \varepsilon)/4}_{\text{negative}} \right\} < 0.$$

It is implied that (23) holds.

(b) If $\Delta s(k) < 0$, by (10), (18), and (24), the proof is similar to (a) and the inequality (23) becomes

$$\Delta s(k) \left\{ \underbrace{\Delta f + \Delta g u_{eq} - \gamma k_1 + \left(\gamma - 1 - \frac{3\alpha}{4}\right)\delta}_{\text{positive}} + \underbrace{(1 - \gamma + 3\beta)\hat{k}_1}_{\text{positive}} + \underbrace{\left(\gamma - 1.5 - \frac{3\alpha}{4}\right)\eta - \frac{3\alpha}{4}\varepsilon}_{\text{positive}} + \underbrace{4s(k) + 4|s(k)|}_{\text{positive}} \right\} < 0,$$

which implies that (23) holds.

(2) when $s(k) < -\varphi(k)$, the proof is similar to (1) and then omitted.

From above all, $\mathbf{B}' = \{e(k) | |s(k)| \leq \varphi(k)\}$ is reachable and stable as the time difference of (19) is negative for $|s(k)| > \varphi(k)$. \square

For all $e(k) \in R^n$, $V(k) \rightarrow \infty$ when $\|e(k)\| \rightarrow \infty$. Thereby, according to the Lyapunov stability theory, s is globally asymptotically stable and converges to \mathbf{B}' .

4 Sliding Mode-Like Fuzzy Logic Control

In this section dynamic fuzzy logic system (DFLS) which includes singleton fuzzier, product reasoning and weight averaging defuzzier will be adopted to implement the SMC presented in Sect. 3. This method will eliminate the chattering.

Consider the same idea in [15], replace the control input u_v with the output of the dynamic fuzzy logic system (DFLS) \bar{u}_v ,

$$\dot{\bar{u}}_v = -\omega_1 \bar{u}_v + \omega_2 \xi^T p \tag{25}$$

where $\omega_1 > 0, \omega_2 > 0$ are filtering parameters to be designed. $\xi = [\xi_1 \dots \xi_m]^T$ is the support points vector of the fuzzy rule base, and $p = [p_1 \dots p_m]^T$ is the fuzzy rule base function vector which is determined by $p_i = \prod_{j=1}^l \mu_j^i / \sum_{i=1}^m \prod_{j=1}^l \mu_j^i$ with l is the total number of input variables, m is the total number of fuzzy reasoning rules, and μ_j^i is the value of the membership function of the j th variable in the i th rule. Suppose the fuzzy partition number of the j th input variable is ζ_j , then the number of fuzzy rules $m = \prod_{j=1}^l \zeta_j$.

Now for convenience, the input variables of DFLS are selected as the sliding mode s and its difference Δs . Their membership functions are all triangular which are shown in Fig. 1a, b, where $\zeta_1 = \zeta_2 = 3$ and thus $m = 9$. The universe field partitions and the membership functions of the output variable are shown in Fig. 1c where $\theta_1, \theta_2 \in R$ are partitioning parameters.

The inferring rules are some IF...THEN...sentences. Their principles are shown in Table 1.

In order to approximate the variable structure control input u_v , the output of the DFLS \bar{u}_v must satisfy the relation between u_v and \hat{k} (the Eq. (7)). So, \hat{k} can be got by

$$\hat{k} = |\bar{u}_v / \hat{g}^{-1} \text{sat}(s, \varphi)| \tag{26}$$

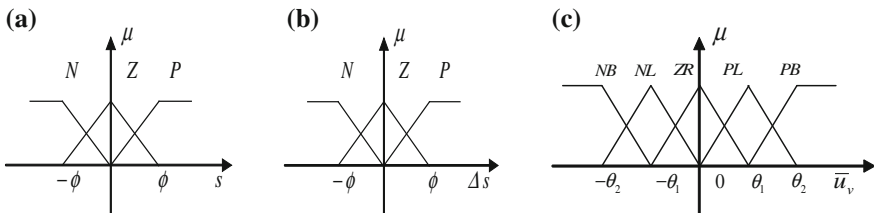


Fig. 1 The fuzzy sets of the variables in the DFLS

Table 1 The fuzzy inferring rules of the DFLS

Rule number	1	2	3	4	5	6	7	8	9
Δs	N	N	N	Z	Z	Z	P	P	P
s	N	Z	P	N	Z	P	N	Z	P
\hat{u}_v	PB	PL	NL	PL	ZR	NL	PL	NL	NB

Then, φ can be self-tuning by the adaptive law (18) while the adaptive law (17) is not used. φ is also the partition parameter of the input variables, so the fuzzy inferring with self-tuning is implemented.

5 Simulation Example

In this section, an example will be given to validate the control design presented in this paper. Simulation results of the example will show the performance of the sliding mode-like fuzzy logic control (SMFC) for discrete nonlinear system.

Consider the system described as

$$x(k + 1) - x(k) = T_s \{ x^3(k) + \exp[x(k)] + \cos x^2(k) + [1 + \cos^2 x(k)]u(k) \} \quad (27)$$

where $x(k)$ is state variable and $u(k)$ is control input. The sampling time T_s is 0.01 second.

For this nonlinear system (27), it is difficult to design a stable controller if the nonlinearity is estimated as $\hat{f} = 1$ and the controller gain $\hat{g} = 1$. But by using our method in this paper, the controller design becomes easier. The sliding mode is designed as $s = x$, and parameters of the adaptive law is preferred as $\alpha = -5$, $\beta = 20$, $\omega_1 = 0.01$, $\omega_2 = 1$, and $\varphi(0) = 0.1$. The DFLS is designed as that proposed in Sect. 4, and the partitions of fuzzy sets, the corresponding memberships, and fuzzy reasoning rules are, respectively, shown in Fig. 1 and Table 1. The parameter $\theta_1 = 2$, $\theta_1 = 50$. The desired trajectory is supposed to be $y_d = 0$, that is, there is no tracked model and signal.

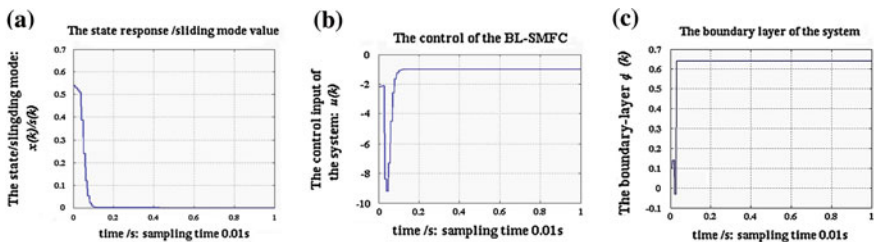


Fig. 2 The simulation results. (a) The system state x . (b) The control input u . (c) The parameter φ

When the initial state $x(0)$ is 0.54, the simulation results of the example are shown in Fig. 2a, b, and c. From these figures, one can see the closed-loop system is stable and the control signal is smooth enough while the boundary of uncertainty is not required during the control design.

6 Conclusions

A sliding mode-like fuzzy logic control (SMFC) is proposed for a class of discrete nonlinear system. Dynamic fuzzy logic system in which parameters are tuned online to approximate the sliding mode control is used, and the overall system stability is analyzed by Lyapunov method. The reachability of the sliding mode is verified, and simulation results validate the proposed control method.

Acknowledgments This work is supported by the Fundamental Research Funds for the Central Universities (No.3142013055), the Science and Technology plan projects of Hebei Provincial Education Department (Z2012089) and Natural Science Foundation of Hebei Province (F2013508110).

References

1. Utkin VI (1977) Variable structure systems with sliding modes. *IEEE Trans Autom Control* 22:212–222
2. Sira-Ramirez H (1989) Nonlinear variable structure systems in sliding mode: the general case. *IEEE Trans Autom Control* 34:1186–1188
3. Edwards C, Spurgeon SK (1998) Sliding mode control: theory and applications. Taylor and Francis, London
4. Utkin VI, Shi JX (1996) Integral sliding mode in systems operating under uncertainty conditions. In: *Proceedings of the 35th conference on decision and control*, vol 4. pp 4591–4596
5. Krupp D, Shtessel YB (1999) Chattering-free sliding mode control with unmodeled dynamics. In: *Proceedings of 1999 American control conference*, vol 1. Arlington, VA, pp 530–534
6. Lee JH, Ko JS (1994) Continuous variable structure controller for BLDDSM position control with prescribed tracking performance. *IEEE Trans Ind Electron* 41:483–491
7. Barrero F et al (2002) Speed control of induction motors using a novel fuzzy sliding mode structure. *IEEE Trans Fuzzy Syst* 10:375–383
8. Wong LK et al (2001) A fuzzy sliding controller for nonlinear systems. *IEEE Trans Ind Electron* 48:32–37
9. Ryu SH, Park JH (2001) Auto-tuning of sliding mode control parameters using fuzzy logic. In: *Proceedings of the American control conference*, vol 1. pp 618–623
10. Lhee CG et al (1999) Sliding-like fuzzy logic control with self-tuning the dead zone parameters. In: *Proceedings of IEEE international fuzzy systems conference*, vol 1. pp 544–549
11. Lhee CG et al (2001) Sliding-like fuzzy logic control with self-tuning the dead zone parameters. *IEEE Trans Fuzzy Syst* 9:343–348

12. Lin WS, Chen CS (2002) Sliding-mode-based direct adaptive fuzzy controller design for a class of uncertain multivariable nonlinear systems. In: Proceedings of the American control conference, vol 3, pp 2955–2960
13. Wai RJ, Lin CM, Hsu CF (2002) Self-organizing fuzzy control for motor-toggle servo mechanism via sliding mode technique. *Fuzzy Sets Syst* 131:235–249
14. Zhang XY, Su HY, Chu J (2003) Adaptive sliding mode-like fuzzy logic control for high-order nonlinear systems. In: Proceedings of the 2003 IEEE international symposium on intelligent control, vol 1, pp 788–792
15. Zhang X (2009) Adaptive sliding mode-like fuzzy logic control for nonlinear systems. *J commun comput* 6(1):53–60

Design of Direct Adaptive Fuzzy Sliding Mode Control for Discrete Nonlinear System

Xiping Zhang and Xiaoyu Zhang

Abstract A new direct adaptive fuzzy sliding mode control (FSMC) design for discrete nonlinear systems is presented to trajectory tracking problem. Firstly, problem formula and dynamic fuzzy logical system (DFLS) are proposed, and then, sliding mode control (SMC) design is constructed based on DFSL in which parameters are self-tuning online. Consequently, the sliding mode is validated using Lyapunov analysis theories that it can be reached by the adaptive law. Thus, the overall system is asymptotically stable and with robustness, chattering free, and adaptive. Finally, the performance of the control design was verified by simulations of an inverted pendulum.

Keywords Discrete nonlinear system · Direct adaptive · Sliding mode control

1 Introduction

For nonlinear systems, sliding mode control (SMC) is an important control method as its robustness, invariance to system uncertainties and resistance to the external disturbance, in which main theory is discussed in [1, 2], and [3]. In recent years, many researchers investigated the application of SMC into adaptive control and intelligent control methods, especially in fuzzy logic control field to overcome the obvious shortcoming of SMC, which is the chattering phenomenon that impedes the application. These are proposed such as [4–6].

X. Zhang (✉) · X. Zhang
Department of Electronics and Information Engineering,
North China Institute of Science and Technology, Beijing, China
e-mail: dxrc1487@ncist.edu.cn

X. Zhang
e-mail: ysuzxy@aliyun.com; ysuzxy@163.com

Fuzzy adaptive control theory is proposed and got great development due to the nonlinear approximation ability of fuzzy logic systems. Many researchers have done some online or offline identification research for nonlinear dynamic function [7–9]. Some researchers apply fuzzy logic systems to SMC to improve the performance of SMC, which accounts for the fuzzy sliding mode control (FSMC, [10–12]). On the other hand, constructing fuzzy logic system by using SMC leads to the sliding mode fuzzy logic control (SMFC, [13–21]). In the FSMC designs, continuous part of SMC controller is obtained by approximating the unknown nonlinear dynamics, and the closed-loop system is verified to be stable based on Lyapunov stability theory. Thereby, adaptive law is found directly which tunes parameters of the fuzzy set supports, that is, rule base functions. C.-C. Chiang and C.C. Hu proposed such a FSMC controller based on feedback linearization [22], and then, Y.C. Hsu and A.M. Heidar studied this method for MIMO nonlinear systems [23]. In other proposed works, one can refer to papers [10–12], etc. For the SMFC designs, there are also many investigations and reports including some direct or indirect adaptive fuzzy control [16] and practical applications [17].

However, numerous of all above researching work considered continuous systems. Papers about the FSMC or SMFC control design for discrete nonlinear systems were sporadically reported.

In this paper, a class of discrete nonlinear system is considered and its FSMC control design is presented. For the discrete nonlinear system which nonlinear dynamic function is unknown but bounded, a dynamic fuzzy logic system (DFLS) is firstly constructed for general nonlinear function approximation. Then, it is used for approximating the unknown dynamics in the SMC controller. The SMC controller is designed to determine that the reach condition of the sliding mode is satisfied, which is validated by the theory of Lyapunov stability theory. Hence, the closed-loop control system is asymptotically stable.

This paper is organized as follows. In Sect. 2, problem formulation is presented. DFSL is introduced in Sect. 3, and main results including adaptive law design and stability validation are reported in Sect. 4. Simulation results of inverted pendulum are proposed in Sect. 5 to verify the controller performance. Conclusions are summarized in Sect. 6.

2 Problem Formula

Considering a class of discrete nonlinear system,

$$\begin{cases} x_i(k+1) = x_{i+1}(k), \\ x_n(k+1) = f(x) + b(x)u(k) + d(x, k), \\ y = x_1(k) \end{cases} \quad (1)$$

where k is the discrete time points, $x = [x_1(k), x_2(k), \dots, x_n(k)]^T$ is state vector, $u(k)$ is control input, nonlinear scalar function $f(x)$, $b(x)$ are unknown but

bounded. Without loss of generality, assuming that $\bar{b}(x) > b(x) > 0$, $\bar{b}(x)$ is its upper bound function. $d(x, k)$ represents the disturbance or unmodeled dynamics.

Define the tracked trajectory as $y_d(k)$, where

$$y_d(k + n) = - \sum_{i=0}^{n-1} a_i y_d(k + i) + r(k), \tag{2}$$

a_0, a_1, \dots, a_{n-1} are Hurwitz polynomial coefficients and $r(k)$ is reference input. Define the error vector,

$$\begin{aligned} e(k) &= [e_1(k), e_2(k), \dots, e_n(k)]^T, \\ e_i(k) &= x_i(k) - y_d(k + i), i = 1, 2, \dots, n. \end{aligned} \tag{3}$$

Then, the control problem is to design fuzzy adaptive sliding mode controller $u(k)$ to make the system tracking error $e(k)$ globally asymptotically stable.

3 Dynamic Fuzzy Logic System

Using singleton fuzzifier, product inference rules, and average defuzzifier method, a DFSL can be described as follows,

$$\hat{g}(x, k + 1) = -\zeta[\hat{g}(x, k) - \Theta^T p] \tag{4}$$

where $g(x, k)$ is nonlinear scalar function to be approximated, $\Theta(k) = [\theta_1(k), \theta_2(k) \dots \theta_m(k)]^T$ is adjustable parameter of the DFSL, $\hat{g}(x, k)$ is the approximation output. $\zeta > 0$ is real scalar to be designed, $p(x) = [p_1(x), p_2(x), \dots, p_m(x)]^T$ is the fuzzy basis function vector, M is the amount of fuzzy rules, H is the state vector dimension.

The following lemma [2–3] is introduced for its application.

Lemma 1 For smooth nonlinear vector field $g(x) : R^n \rightarrow R$, there is a parameter

$$\Theta^* = \arg \min_{\theta \in \Omega_\theta} \left[\sup_{x \in R^n} |g(x) - g^*(x)| \right],$$

to guarantee $\forall \varepsilon \in R, \varepsilon > 0, \|g(x) - g^*(x)\| < \varepsilon$, where $g^*(x)$ is the approximation output of fuzzy logic system of DFSL (4), namely $g^*(x) = \Theta^{*T} p$. The fuzzy inference rules of the fuzzy logic system are designed as follows.

- If $x_1(k)$ is a_1^1 and $x_2(k)$ is a_2^1 and $\dots \dots x_H(k)$ is a_H^1 then $\hat{g}(x)$ is θ_1
- If $x_1(k)$ is a_1^2 and $x_2(k)$ is a_2^2 and $\dots \dots x_H(k)$ is a_H^2 then $\hat{g}(x)$ is θ_2
- $\dots \dots$
- If $x_1(k)$ is a_1^M and $x_2(k)$ is a_2^M and $\dots \dots x_H(k)$ is a_H^M then $\hat{g}(x)$ is θ_M

In the SMC controller design, the DFSL described by (4) and the above rules of the fuzzy logic system are adopted directly for the approximation of unknown dynamics.

4 Main Results

4.1 Adaptive Sliding Mode Control Law Design

A predefined sliding mode surface

$$S(k) = c_n e_n(k) + \sum_{i=1}^{n-1} c_i e_i(k) \tag{5}$$

where $c_i > 0 (i = 1, \dots, n - 1)$ are Hurwitz polynomial coefficients and $c_n > 0$.

In order to carry out the subsequent control law design, the following lemma [24] is introduced.

Lemma 2 *For Hurwitz polynomial parameters $c_i > 0 (i = 1, \dots, n - 1)$ and $c_n > 0$, there are always parameters $\lambda_i > 0 (i = 0, 1, 2, \dots, n)$ which guarantee*

$$D = \begin{bmatrix} \lambda_0 & 0 & \cdots & 0 & c_1 \\ 0 & \lambda_1 & \cdots & 0 & c_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_n & c_n \\ c_1 & c_2 & \cdots & c_n & \lambda_n \end{bmatrix} > 0, D = D^T.$$

For the convenience of the design, we define

$$S_a(k) = \frac{1}{2} \sum_{i=0}^{n-1} \lambda_i [e_{i+1}^2(k) - e_i^2(k)] - \frac{\lambda_n}{2} e_n^2(k) - s(k - 1)e_n(k),$$

where $e_0(k) = e_1(k - 1)$. And define a nonlinear dynamic function to be approximated as,

$$g(k) = \frac{S_a + 0.5\lambda_n[f(x) + b(x)u(k) + d(x, k)]^2}{S\bar{b}(x)} + \frac{f(x) + [b(x) - \bar{b}(x)]u(k) + d(x, k)}{\bar{b}(x)}. \tag{6}$$

Consequently, a sliding mode controller is designed as

$$\begin{aligned}
 u(k) = & -\bar{b}^{-1}(x) \sum_{i=0}^{n-1} a_i y_d(k) + \bar{b}^{-1}(x) r(k) \\
 & - \hat{g}(k) - \{ \varepsilon_h + [k_1 \bar{b}^{-1}(x) + k_2 \bar{b}(x)] \|S\| \} \text{sgn}S,
 \end{aligned}
 \tag{7}$$

which parameters

$$k_1 > 0, k_2 > \frac{\alpha^2}{2} + \frac{\alpha}{2} (1 - \beta)^2 \|G\|^{-1},
 \tag{8}$$

and the adaptive laws of the adopted DFSL are designed as

$$\begin{cases}
 \Delta\Theta(k) = (1 - \beta)[G^{-1}]^T p S \bar{b}(x) \\
 \Delta\hat{g}(k) = -\beta[\hat{g}(k) - \Theta^T P] + [\alpha + (1 - \beta)p^T G^{-1} p] S \bar{b}(x)
 \end{cases}
 \tag{9}$$

where $G \in R^{M \times M}$, $G = G^T$, $G > 0$ is optional parameter matrix, $\alpha > 0$, $0 < \beta < 1$ are optional scalar parameters, $\varepsilon_h > |\varepsilon|$ with $\varepsilon = g(x) - g^*(x)$ is minimum approximation error. According to Lemma 1, approximation error will be arbitrarily small, and then, ε_h can be selected as any size of real according to the situation of system uncertainties and disturbances.

The designed fuzzy adaptive sliding mode controller consists of four parts which are DAFLS approximator $\hat{g}(k)$, adaptive mechanism, sliding mode controller, and reference input.

In Eq. (9), the first line describes adaptive mechanism, while the second line describes DAFLS, from which we can see that both the operation of DAFLS and the adaptive mechanism need the sliding mode value $S(k)$. When $S(k) = 0$, adaptive mechanism is in a steady state. The SMC controller (4, 7–9) constitutes a complex dynamic subsystem. In the following, the analysis of the closed-loop system stability is carried through.

4.2 Stability Analysis

Main result of system stability is summarized to be the theorem as follows.

Theorem 1 *For nonlinear system (1) and the given reference trajectory (2), the sliding mode (5) is reachable and the system tracking error is globally asymptotically stable under the FSMC control (7–8) which is based on the DFSL (4, 9) if the coefficients of sliding mode (5) $c_i > 0$ ($i = 1, 2, \dots, n - 1$) is Hurwitz.*

Proof Choose the Lyapunov function of the tracking error as

$$V(k) = \frac{\gamma}{2} \tilde{e}^T(k) D \tilde{e}^T(k) + \frac{1}{2} [\hat{g} - \Theta^T p]^2 + \frac{\alpha}{2} \tilde{\Theta}^T G \tilde{\Theta},
 \tag{10}$$

where $\tilde{e}(k) = [e_1(k - 1), e_1(k), \dots, e_n(k)]^T$, $\gamma = \alpha(1 - \beta)$, $\tilde{\Theta} = \Theta^* - \Theta$ is the error between Θ and optimum parameter Θ^* . According to Lemma 2, the matrix $D = D^T$, $D > 0$ and hence obviously $V > 0$. Furthermore, when $\|e\| \rightarrow \infty$, $\|V(k)\| \rightarrow \infty$.

Seek the time difference of Lyapunov function (10),

$$\begin{aligned} \Delta V(k) &= \gamma S_a(k) + \gamma S(k)e_n(k + 1) + \frac{\gamma \lambda_n}{2} e_n^2(k + 1) \\ &\quad + \frac{\alpha}{2} [\Theta^* - \Theta(k + 1)]^T G [\Theta^* - \Theta(k + 1)] + \frac{1}{2} [\hat{g}(k + 1) - \Theta^T(k + 1)p]^2 \\ &\quad - \frac{1}{2} [\hat{g}(k) - \Theta^T(k)p]^2 - \frac{\alpha}{2} [\Theta^* - \Theta(k)]^T G [\Theta^* - \Theta(k)] \end{aligned}$$

Then, substitute the adaptive law (9) into the above equation, one has

$$\begin{aligned} \Delta V(k) &= \gamma S_a(k) + \gamma S(k)e_n(k + 1) + \frac{\gamma \lambda_n}{2} e_n^2(k + 1) \\ &\quad + \alpha(1 - \beta) S \bar{b}(x) [\hat{g}(k) - \Theta^T(k)p] - \alpha(1 - \beta) \hat{g}^*(k) S \bar{b}(x) \\ &\quad + \alpha(1 - \beta) S \bar{b}(x) \Theta^T(k)p + \frac{\alpha}{2} (1 - \beta)^2 p^T G^{-1} p S^2 \bar{b}^2(x) \\ &\quad + \frac{1}{2} \alpha^2 S^2 \bar{b}^2(x) + \left(\frac{\beta^2}{2} - \beta \right) [\hat{g}(k) - \Theta^T(k)p]^2. \end{aligned}$$

Because $0 < \beta < 1$, the above equation satisfies that

$$\begin{aligned} \Delta V(k) &\leq \gamma S_a(k) - \bar{b}(x) [\hat{g}^*(k) - \hat{g}(k)] + \frac{\gamma \lambda_n}{2} e_n^2(k + 1) + \frac{1}{2} \alpha^2 S^2 \bar{b}^2(x) \\ &\quad + \gamma S(k)e_n(k + 1) + \frac{\alpha}{2} (1 - \beta)^2 p^T G^{-1} p S^2 \bar{b}^2(x) \end{aligned}$$

and then using the SMC controller (7),

$$\begin{aligned} \Delta V(k) &\leq \gamma S \bar{b}(x) [g(k) - \hat{g}(k) - \{ \varepsilon_h + [k_1 \bar{b}^{-1}(x) + k_2 \bar{b}(x)] \|S\| \} \text{sgn}S] \\ &\quad - \gamma S \bar{b}(x) [\hat{g}^*(k) - \hat{g}(k)] + \frac{1}{2} \alpha^2 S^2 \bar{b}^2(x) + \frac{\alpha}{2} (1 - \beta)^2 p^T G^{-1} p S^2 \bar{b}^2(x). \end{aligned}$$

It is known that p is the fuzzy basis, hence $p^T p \leq 1$ and then $p^T G^{-1} p \leq \|G\|^{-1}$.

According to Eq. (8), $\Delta V(k) \leq -\gamma k_1 \|S\|^2$, $\Delta V(k)$ is semi-negative definite.

$\Delta V(k) \equiv 0$ holds if and only if $\|S(k)\| \equiv 0$. $\forall e \in R^n$, $\|S(k)\| \neq 0$, then $\Delta V(k) < 0$. According to the Lyapunov stability theory, the system error asymptotically converges to $\Omega_e \stackrel{\text{Def}}{=} \{e(k) \mid \|S(k)\| = 0\}$, namely the sliding mode can be reached.

After the system state reaches to the sliding mode, that is, $S(k) = 0$, the system error equation becomes the following $n - 1$ order system described as

$$\begin{cases} e_i(k + 1) = e_{i+1}(k), & i = 1, 2, \dots, n - 2 \\ c_n e_{n-1}(k + 1) = -c_1 e_1(k) - c_2 e_2(k) - \dots - c_{n-1} e_{n-1}(k). \end{cases}$$

If $c_i > 0 (i = 1, 2, \dots, n - 1)$ is Hurwitz, then obviously the above $n - 1$ order system is asymptotically stable. As a result, the overall closed-loop system is asymptotically stable and the theorem is proved.

5 Experiment on Inverted Pendulum System

Inverted pendulum systems are typically multi-coupling nonlinear systems. Figure 1 shows a schematic diagram of an inverted pendulum experiment system in which the dolly is driven by a servo motor and can move in the horizontal direction. If the friction resistance f_0 during the inverted pendulum horizontal movement and shaft resistance to rotation f_1 are all ignored, the discrete mathematical model is described as

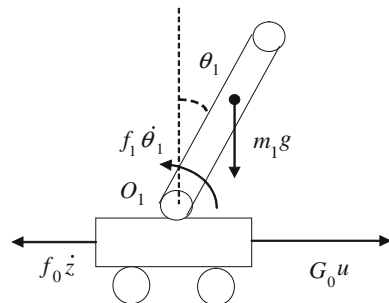
$$\begin{aligned} x_1(k + 1) &= x_2(k) \\ x_2(k + 1) &= f(x) + b(x)u(k), \end{aligned} \tag{11}$$

when only the swing angle is considered as the state of the system $x_k = [\theta_k, \Delta\theta_k]^T$.

The estimation of upper bound for $b(x)$ was $\bar{b} = 500$. The reference model and signal were all designed as zero. The sliding mode $S(k) = x_2(k) + 0.05x_1(k)$. The FSMC controller was designed as (7, 8), and parameters of its adaptive law (9) were designed as $\varepsilon_h = 0.05, \kappa = 0.6, \beta = 0.95, G = I$ (the identity matrix). Membership function for DFLS was selected as Gauss function. Based on all above control design, the application experiment on inverted pendulum was executed.

The initial angle excursion was $\theta(0) = \pi/50$. And the subsequent angular excursion curve is shown in Fig. 2. The corresponding control volt curve of the dolly is shown in Fig. 3. During this experimental process, the value of the sliding mode $S(k)$ is drawn in Fig. 4.

Fig. 1 Schematic diagram of an inverted pendulum



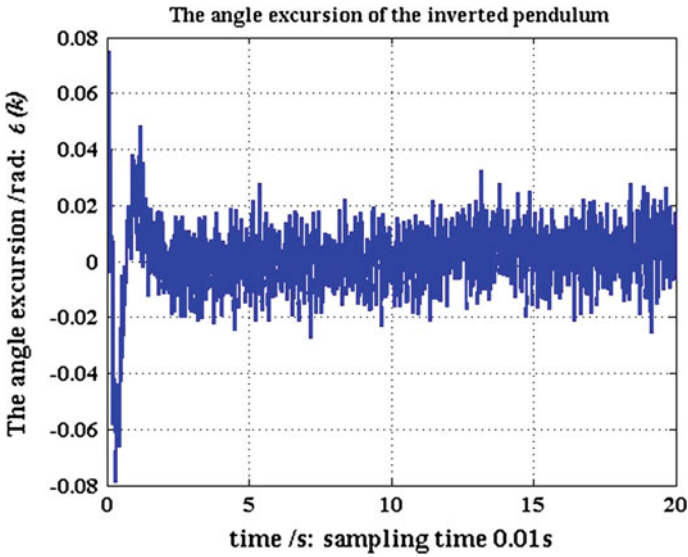


Fig. 2 The angle excursion curve

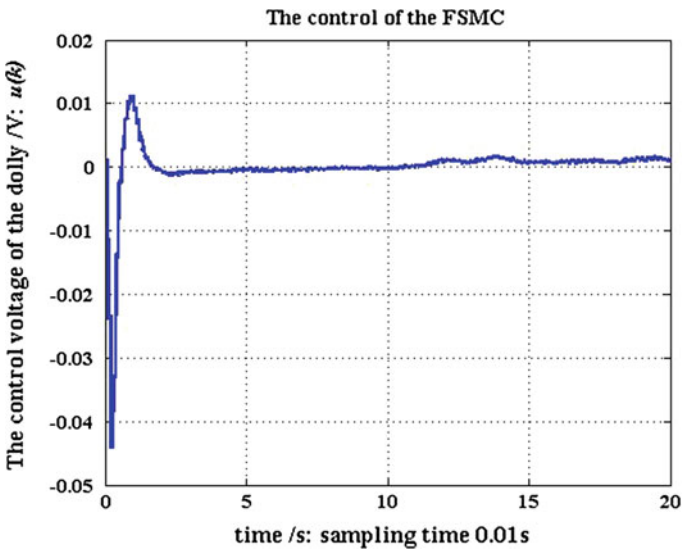


Fig. 3 The control volt curve

In order to validate the robustness of the control design, disturbance was added into the experiment system through the pendulum. Figure 5 shows that the angle excursion curve is still stable while the disturbance exists. The pendulum angle converges to the origin although the steady state was interfered.

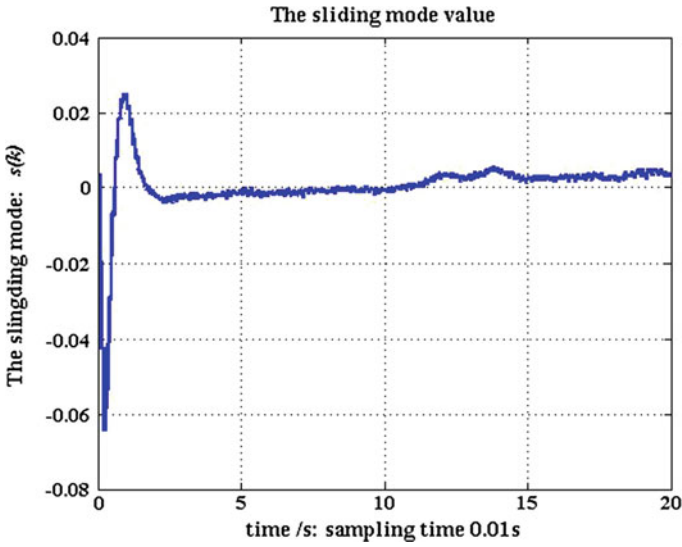


Fig. 4 The value curve of the sliding mode

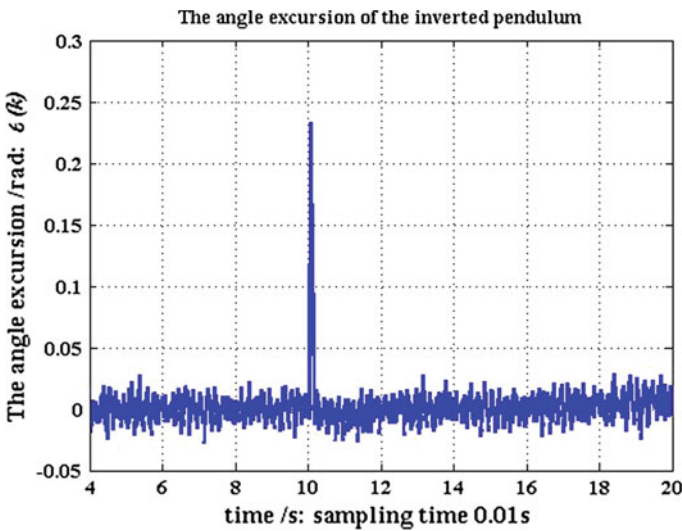


Fig. 5 The control volt curve when the system is disturbed

6 Conclusions

A FSMC control design method is proposed in this paper based on DFSL, which has showed strong arbitrary approximation properties to unmodeled dynamics. And furthermore, DFSL has filtering function in the presented SMC design; thus,

the chattering of the SMC is greatly weakened while the robustness is reserved. Finally, its application simulation results to inverted pendulum system validated the performance of the proposed method.

Acknowledgments This work is supported by the Fundamental Research Funds for the Central Universities (No.3142013055), the Science and Technology plan projects of Hebei Provincial Education Department (Z2012089) and Natural Science Foundation of Hebei Province (F2013508110).

References

1. Utkin VI (1977) Variable structure systems with sliding modes. *IEEE Trans Autom Control* 22:212–222
2. Sira Ramirez H (1989) Nonlinear variable structure systems in sliding mode: the general case. *IEEE Trans Autom Control* 34:1186–1188
3. Edwards C, Spurgeon SK (1998) Sliding mode control: theory and applications. Taylor & Francis, London (UK)
4. Morioka H et al (1995) Neural network based chattering free sliding mode control. In: Proceedings of the 34th SICE annual conference (International session papers), vol 2, pp 1303–1308
5. Zhang DQ, Panda SK (1999) Chattering-free and fast response sliding mode controller. *IEE Proc-D: Theor Appl* 146:171–177
6. Utkin VI, Shi JX (1996) Integral sliding mode in systems operating under uncertainty conditions. In: Proceedings of the 35th conference on decision and control, vol 4, pp 4591–4596
7. Wang S, Yu D (2000) Error analysis in nonlinear system identification using fuzzy system. *J Software* 11:447–452
8. Lee JX, Vukovich G (1997) Identification of nonlinear dynamic systems—a fuzzy logic approach and experimental demonstrations. In: Proceedings IEEE international conference systems, man, and cybernetics, Orlando, Florida, pp 1121–1126
9. Wang L (1995) Design and analysis of fuzzy identifiers of nonlinear dynamic systems. *IEEE Trans Aut Contr* 40:11–23
10. Temeltas H (1998) A fuzzy adaptation technique for sliding mode controllers. In: Proceedings IEEE international symposium on industrial electronics, Pretoria, South Africa, pp 110–115
11. Barrero F et al (2002) Speed control of induction motors using a novel fuzzy sliding mode structure. *IEEE Trans Fuzzy Syst* 10:375–383
12. Wong LK et al (2001) A fuzzy sliding controller for nonlinear systems. *IEEE Trans Industr Electron* 48:32–37
13. Chen JY, Lin YH (1995) A self-tuning fuzzy controller design. In: Proceeding of IEEE international conference on neural networks, vol 3, pp 1358–1362
14. Palm R (1992) Sliding mode fuzzy control. In: Proceedings of IEEE international conference on fuzzy systems, vol 1, pp 519–526
15. Ryu SH, Park JH (2001) Auto-tuning of sliding mode control parameters using fuzzy logic. In: Proceedings of the American control conference, vol 1, pp 618–623
16. Tu KY, Lee TT, Wang WJ (2000) Design of a multi-layer fuzzy logic controller for multi-input multi-output systems. *Fuzzy Sets Syst* 111:199–214
17. Wai RJ, Lin CM, Hsu CF (2002) Self-organizing fuzzy control for motor-toggle servomechanism via sliding mode technique. *Fuzzy Sets Syst* 131:235–249

18. Lin HR, Wang WJ (1998) Fuzzy control design for the pre-specified tracking with sliding mode. In: Proceedings IEEE World congress on computational intelligence, Anchorage, Alaska, pp 292–295
19. Lhee C-G, Park J-S, Ahn H-S, Kim D-H (2001) Sliding mode-like fuzzy logic control with self-tuning the dead zone parameters. *IEEE Trans Fuzzy Syst* 9:343–348
20. Zhang XY, Su HY, Chu J (2003) Adaptive sliding mode-like fuzzy logic control for high-order nonlinear systems. In: Proceedings of the 2003 IEEE international symposium on intelligent control, vol 1, pp 788–792
21. Zhang X (2009) Adaptive sliding mode-like fuzzy logic control for nonlinear systems. *J Commun Comput* 6:53–60
22. Chiang CC, Hu CC (1999) Adaptive fuzzy controller for nonlinear uncertain systems. In: Proceedings of the second international conference on intelligent processing and manufacturing of materials, vol 2, pp 1131–1136
23. Hsu C, Heidar AM (1998) Fuzzy variable structure control for MIMO systems. In: Fuzzy systems proceedings, IEEE world congress on computational intelligence, vol 1, pp 280–285
24. Zhang X, Su H (2004) Sliding mode variable structure state norm control of SISO linear systems. *Control Eng China* 11:413–418

Application of Improved Particle Swarm Optimization-Neural Network in Long-Term Load Forecasting

Xuelian Gao, Yanyu Chen, Zhennan Cui, Nan Feng, Xiaoyu Zhang and Lei Zhao

Abstract The improved particle swarm optimization-neural network (IPSO-NN) can be achieved by improving four aspects of the classical particle swarm optimization (CPSO), such as the inertia weight, the learning factor, the variation factor, and objective function. By applying CPSO, the neural network (NN), and IPSO-NN into the long-term power load forecasting problem, the results show that IPSO-NN has not only better global searching ability and higher convergent accuracy than CPSO does but also shorter training time and faster convergent speed than NN does. In feasible running time, IPSO-NN owns the smallest mean error and the acceptable relative error within 3 %. Finally, this paper applies IPSO-NN in the long-term load forecasting of Langfang city from 2010 to 2019.

Keywords Power load forecasting · Improved particle swarm-neural network (IPSO-NN) · Long-term load forecasting

1 Introduction

It is important to draw up a plan for the development of the power system. The basis of the plan is the long-term load forecasting that also provides the macro decision-making basis for the reasonable arrangement for the construction of power plants and power grids. The forecasting aims at how to make the power

Supported by “the Fundamental Research Funds for the Central Universities”

X. Gao (✉) · Y. Chen · Z. Cui · N. Feng · X. Zhang · L. Zhao
School of Electrical and Electronic Engineering, North China Electric Power University,
Beijing 102206, China
e-mail: xuelian_gao@ncepu.edu.cn

construction to meet the growth of national economy and the improvement of living standards [1]. But the historical data of the long-term power load forecasting are relatively short, and some of the data are almost meaningless. Furthermore, the load forecasting is also influenced by two factors such as economic and social development. All those features make an accurate forecast of the long-term power load become very difficult.

The research shows that the data of long-term power load forecasting increase monotonically year by year, but vary with uncertainty [2]. Classical particle swarm optimization (CPSO) is generally applied to the power load forecasting by using polynomial or exponential function as the objective function [3, 4]. On the other side, long-term power load is not entirely in accordance with the elementary functions but usually with the characteristics of uncertainty and randomness. So, the elementary function acts as the objective function, which cannot meet the accuracy requirements of the load forecasting largely. Many scholars have proposed several improved CPSOs which are already applied to the power load forecasting problem [5–7]. Such as papers [8] and [9] combine the particle swarm optimization with a neural network (NN) which is applied to the short-term power load forecasting problem with a reasonable solution.

This paper introduces IPSO-NN by combining the improved particle swarm optimization with the neural networks. IPSO-NN uses NN as the objective function not only avoids the defects of the objective function, ensures the requirement of result precision and global convergence, but also owns the nonlinear and uncertain characteristics to fit the long-term power load forecasting. Finally, the comparisons of simulation among IPSO-NN, NN, and CPSO are provided to validate that IPSO-NN calculates faster than NN and operates more accurate than CPSO. The arithmetic error is controlled within 3 %, which makes IPSO-NN effective for the long-term load forecasting of the power system.

2 CPSO

Assuming that the size of the particle groups is N , and the position of particle i ($i = 1 \sim N$) in the M -dimensional search space can be expressed as $p_i = (p_{i1}, p_{i2}, \dots, p_{im}, \dots, p_{iM})$. The flying velocity of the particles is described by $v_i = (v_{i1}, v_{i2}, \dots, v_{im}, \dots, v_{iM})$, which means a particle's moving distance in each iteration. CPSO can be expressed by (1–2).

$$v_{im}^{(k+1)} = \omega \cdot v_{im}^{(k)} + c_1 \cdot r_1^{(k)} \left(p_{ipm}^{(k)} - p_{im}^{(k)} \right) + c_2 \cdot r_2^{(k)} \left(p_{gm}^{(k)} - p_{im}^{(k)} \right) \quad (1)$$

$$p_{im}^{(k+1)} = p_{im}^{(k)} + v_{im}^{(k+1)} \quad (2)$$

p_{ipm} , individual optimum value in (1), is the best historical position of the current particle. The distance between p_{ipm} and the current position of the particle p_{im} is used to set the direction of particle random motion. p_{gm} , group optimum value in (1), is the best historical position for the entire population of particles. The distance between p_{gm} and p_{im} is used to change the incremental motion of the current particle toward the group optimum value. c_1 and c_2 are learning factors. ω is the inertia weight.

CPSO is simple and easy to realize with the characteristic of fast convergence speed, but the algorithm is easy to jump into local optimum value.

3 IPSO-NN

The improved CPSO named IPSO-NN in this paper improves the inertia weight and learning factors, adds the variation items, and uses the NN as its objective function. Taking into account all the four aspects of the improvement, IPSO-NN can be described by (3-7).

$$v_{im}^{(k+1)} = \omega^{(k)} \cdot v_{im}^{(k)} + c_1^{(k)} \cdot r_1^{(k)} \left(p_{ipm}^{(k)} - p_{im}^{(k)} \right) + c_2^{(k)} \cdot r_2^{(k)} \left(p_{gm}^{(k)} - p_{im}^{(k)} \right) \tag{3}$$

$$c_1^{(k)} = a - \lambda \cdot k \tag{4}$$

$$c_2^{(k)} = b + \lambda \cdot k \tag{5}$$

$$\omega^{(k)} = c - \lambda \cdot k \tag{6}$$

$$p_{im}^{(k+1)} = p_{im}^{(k)} + v_{im}^{(k+1)} + f \cdot r_3^{(k)} \tag{7}$$

c_1 , c_2 , and ω change linearly with the iterations. In consideration of the convergence and accuracy of the algorithm, the related parameters can be determined by experiences: $a = 0$, $b = 1.9$, $c = 0.8$, $\lambda = 10^{-4}$.

3.1 Learning Factor Improvement

Learning factor c_1 , showed in (4), controls the influence of “self-learning” section of the particle velocity. Learning factor c_2 , showed in (5), controls the influence of “social learning” section of the particle velocity. By changing the value of c_1 and c_2 dynamically, “self-learning” part believes themselves more and accounts for a slightly larger proportion in the early stage of iterations, while “social learning” part takes into account “social” global optimum more and accounts for a slightly

larger proportion in the later stage of iterations. This is conducive to the algorithm convergence in the global optimum and to the accuracy of the algorithm convergence.

3.2 Inertia Weight Improvement

The inertia weight in CPSO is to control the influence that the particle speed in previous iteration has an impact on the particle speed of the current iteration. Larger inertia weight can strengthen the global search ability of the particle swarm algorithm, while smaller inertia weight can enhance the local search ability.

In this paper, the inertia weight ω , showed in (6), decreases linearly. This method can make the particle group to explore greater area and gain the optimal solution faster in the early stage. With the increase in iteration times, ω will reduce little by little and the particle speed will slow down, which help the IPSO-NN finish local search job better. In short, the inertia weight ω improves the global search ability as well as the convergence speed of the algorithm.

3.3 Mutation Operation

In order to avoid local convergence, this paper combines CPSO with the mutation operation in genetic algorithm. The particle position in CPSO is considered as an individual. When conditions are fulfilled, the variation is executed.

As shown in (7), a smaller random number r_3 , named the disturbance of variation, is added to increase the diversity of the individual, so that the particles can break the original aggregation state and get out of the local optimal position to avoid converging too fast. f is a flag bit. When f meets the variable condition, its value is 1, otherwise is set to zero.

In order to determine the variable condition, this paper uses (8–11) to evaluate the solution.

3.3.1 Fitness Function

f_i is the fitness value of the current particle i ; g_{ij} is the power load of year j corresponding to the current particle i ; g_{0j} is the real power load of year j .

$$f_i = \frac{1}{\frac{1}{2} \sum_{j=1}^n (g_{ij} - g_{0j})^2} \quad (8)$$

3.3.2 Average Fitness Function

$$f_{\text{avg}} = \frac{1}{N} \sum_{i=1}^N f_i \quad (9)$$

f_{avg} is the average fitness function. N is the scale of the particle swarm.

3.3.3 Group Fitness Variance

$$\sigma^2 = \sum_{i=1}^N (f_i - f_{\text{avg}})^2 \quad (10)$$

σ^2 is the group fitness variance of particle swarm reflecting the convergent degree of all the particles in the swarm. Smaller the σ^2 is, more convergent the particle swarm is [10].

3.3.4 Variation Condition

$$\begin{cases} \sigma^2 < \varepsilon \\ f_g > f_i \end{cases} \quad (11)$$

The fitness variance decides the timing of variation that is operated when the condition of (11) is satisfied. ε is the convergent precision, which value is determined by the actual situation. f_i is the optimal theoretical value or the optimal empirical value of the group fitness of the particle swarm.

3.4 NN Objective Function

The objective functions of CPSO commonly use polynomial model, exponential model, or G (1,1) model [11], which approximate the load curve into a specific function. Load forecasting is affected by so many factors that the load curve is carried out strictly in accordance with the specific function by no means. This paper uses the NN as the objective function of IPSO-NN, because the NN has the characteristics of nonlinearity and self-adaption to fit the load curve better. The connection weights of NN are the target parameters which the IPSO-NN aims to optimize.

4 Steps of IPSO-NN

The steps of IPSO-NN are shown in Fig. 1.

1. Initialize IPSO-NN.
2. Take the value of the current particle into the NN function and get the output value (result of power load forecasting).
3. Calculate the fitness value of each particle, as well as the average fitness and group fitness variance based on the output value and fitness.

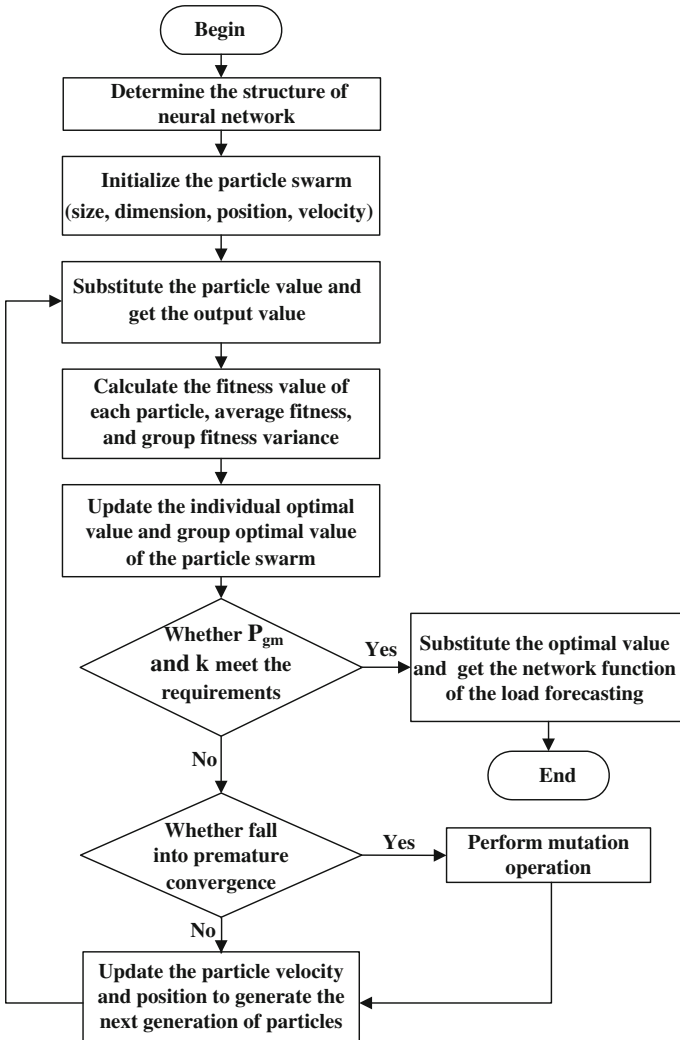


Fig. 1 Flow diagram of IPSO-NN

4. Update p_{ipm} and p_{gm} .
5. Determine whether the value of p_{gm} meets the requirements or the number of iterations meets the maximum iterations k . If either condition is fit, the algorithm ends or continues to step 7.
6. Determine whether the algorithm falls into premature convergence. If happens, act the mutation operation or skip to next step.
7. Update the velocity and position to generate the next generation of particles, and then repeat the steps of 3–6.

5 Example Analysis

Statistical database of China Economic Information Network provides the power load of Langfang city from 1998 to 2009 (as shown in Table 1). Based on the reference data, IPSO-NN predicts the long-term power load of Langfang.

5.1 Algorithm Comparison

The construct of IPSO-NN is a 3-layer NN with 4 inputs, 15 hidden layer nodes, and 1 output. The NN owns 75 connection weights that are optimized by the forecasting objective function. Initial parameters of IPSO-NN are set, respectively: the size of the particle swarm $N = 20$, the dimensions of particle swarm $M = 75$, maximum iterations $k = 5,000$.

Using of rolling forecasted method, the power load data of Langfang city for 12 years shown in Table 1 are divided into 7 groups. Each group regards the data of previous four years as inputs and the data of fifth year as the output. Then, these seven groups of data are the training samples of NN.

Figure 2 presents the prediction results of IPSO-NN, and Fig. 3 shows the relative error between forecasting results and real value.

Table 1 Power load data of Langfang city

Years	Annual power generation (million kilowatt hours)	Years	Annual power generation (million kilowatt hours)
1998	70,216	2004	115,841
1999	78,555	2005	147,977
2000	83,028	2006	174,602
2001	85,014	2007	214,499
2002	90,850	2008	248,072
2003	100,140	2009	295,809

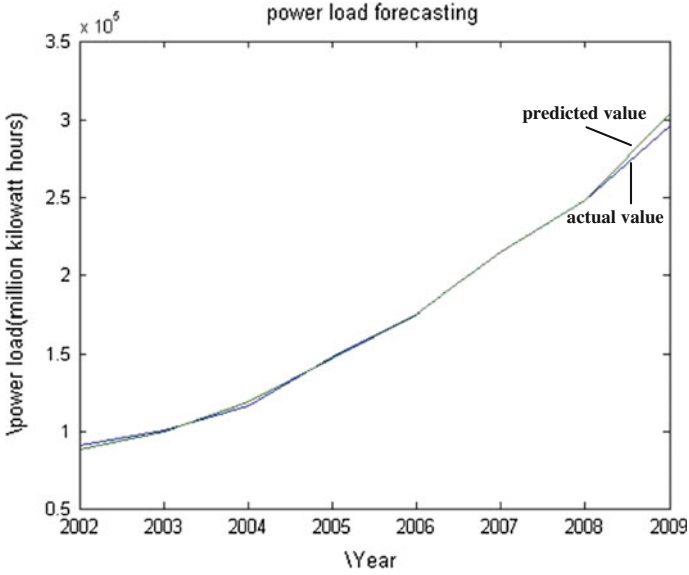


Fig. 2 Prediction curve of IPSO-NN

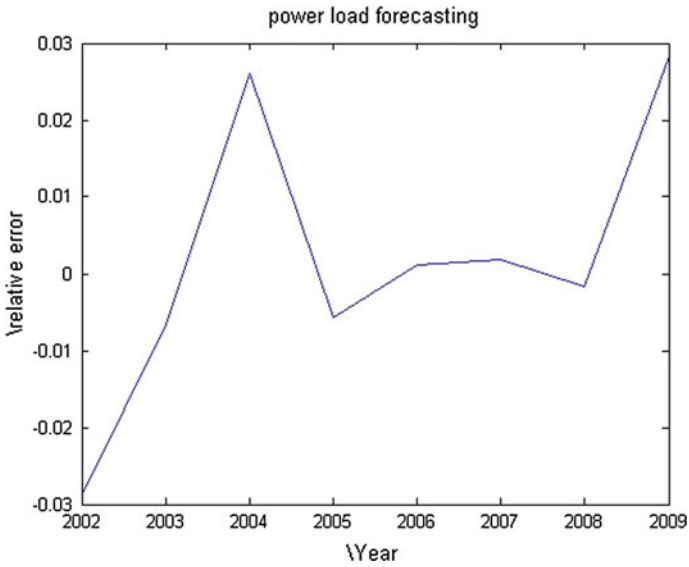


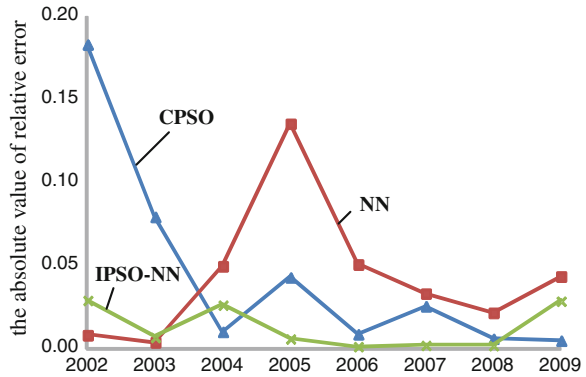
Fig. 3 Relative error of IPSO-NN

With the same data and software, this paper adopts CPSO, NN, and IPSO-NN to forecast long-term power load of Langfang city, respectively. CPSO and IPSO-NN share the same population size, and the objective function of CPSO is

Table 2 Comparison among CPSO, NN, and IPSO-NN

Algorithm	Iterations	Running time(s)	Average error
CPSO	1,750	27	0.17
NN	5,000	106	0.15
IPSO-NN	1,300	76	0.05

Fig. 4 Comparison of power load forecast error



$$y = a \cdot e^{bx} + c \cdot x^2 + d \cdot x + e \tag{12}$$

NN with the structure of 4 inputs, 15 hidden layer nodes, and 1 output is trained by the gradient descent method. Table 2 shows the comparison among CPSO, NN, and IPSO-NN, which contains items of iterative numbers, the running times, and the average error. Figure 4 shows the error curves of the three prediction methods from 2002 to 2009.

From Table 2 and Fig. 4, the iterative number of CPSO is smaller than NN’s but larger than IPSO-NN’s. Meanwhile CPSO has the least running time, but its average error is worst. The average error of NN, whose iterative number and running time is much larger than the other two, is slightly less than CPSO. Taken together, the running time of IPSO-NN is between the other two algorithms due to the introduction of the NN as the objective function. Compared to the CPSO, the objective function of IPSO-NN is much more complex which results in longer running time, but the accuracy of IPSO-NN is much higher than the CPSO.

Table 2 and Fig. 4 can prove that IPSO-NN combines the global search ability of NN with the fast convergent feature of CPSO. IPSO-NN not only reduces the computation time and the average error significantly but also improves the accuracy.

Table 3 Predicted power load of Langfang city over the next ten years

Years	Load forecast (million kilowatt hours)	Years	Load forecast (million kilowatt hours)
2010	330,260	2015	475,249
2011	377,633	2016	480,095
2012	399,597	2017	482,823
2013	439,701	2018	485,089
2014	455,286	2019	488,589

5.2 Application of IPSO-NN

The forecasting power load results of Langfang city in the next 10 years are given in Table 3. Due to the training samples have little influence on load forecasting for years far away, the load forecasting values of 2016–2019 are stable. We can conclude that IPSO-NN is more suitable for medium- and long-term power load forecasting.

6 Conclusion

By improving the inertia weight and learning factor of CPSO, adding the mutation operation in genetic algorithm, and applying NN to the objective function, IPSO-NN not only overcomes the shortcomings of CPSO in premature and in low convergence precision but also results the problems of NN in long training time and in slow convergence. The comparison among IPSO-NN, CPSO, and NN shows that IPSO-NN can be applied to long-term load forecasting of power system effectively by its stable output, fast convergence, and forecast error within 3 %. Finally, the paper takes the power load forecasting of Langfang city as an example, applying IPSO-NN to the long-term power load forecasting and obtaining reliable results.

References

1. Kang CQ, Xia Q (2007) Load forecast of power system. China Electric Power Press, Beijing
2. Fang RC, Zhou JZ (2008) Application of particle swarm optimization based nonlinear grey Bernoulli model in medium- and long- term load forecasting. *Power Syst Technol* 32(12):60–63 (in Chinese)
3. Zhang YD, Wu LN et al (2009) Chaotic immune particle swarm optimization for multi-exponential fitting. *J SE Univ (Nat Sci Ed)* 39(4):678–683 (in Chinese)
4. Duan Q, Zhao JG et al (2010) Relevance vector machine based on particle swarm optimization of compounding kernels in electricity load forecasting. *Electr Machin Control* 14(6):33–38 (in Chinese)
5. Zhang ZY (2006) Particle swarm optimization algorithm and its application in the optimal operation of power system. Tianjin University, Tianjin

6. Wu CY, Wang FL et al (2009) Application of improved particle swarm optimization in power load combination forecasting model. *Power Syst Technol* 33(2):27–30 (in Chinese)
7. Ding YF (2005) Particle swarm optimization algorithm and its applications to power systems economic operation. Huazhong University of Science and Technology, Wuhan
8. Shi B, Li YX et al (2009) Short-term load forecasting based on modified particle swarm optimization and radial basis function neural network model. *Power Syst Technol* 33(17):180–184 (in Chinese)
9. Liu L, Yan DJ, et al. (2006) New method for short term load forecasting based on particle swarm optimization and fuzzy neural network. In: *Proceedings of the CSU-EPSCA*, vol 18(3), pp. 40–43. (in Chinese)
10. Wang CR, Duan XD et al (2004) A modified basic particle swarm optimization algorithm. *Comput Eng* 30(21):35–37 (in Chinese)
11. Cao GJ, Huang C et al (2004) Load forecasting based on improved GM(1,1) model. *Power Syst Technol* 28(13):50–53 (in Chinese)

Routing Techniques Based on Swarm Intelligence

Delfín Rupérez Cañas, Ana Lucila Sandoval Orozco
and Luis Javier García Villalba

Abstract Artificial immune systems (AISs) are used for solving complex optimization problems and can be applied to the detection of misbehaviors, such as fault tolerance. We present novel techniques for the routing optimization from the perspective of the artificial immunology theory. We discussed the bioinspired protocol AntOR and analyzed its new enhancements. This ACO protocol based on swarm intelligence takes into account the behavior of the ants at the time of obtaining the food. In the simulation results, we compare it with the reactive protocol AODV, observing how our proposal improves it according to the delivered data packet ratio and overhead in number of packet metrics.

Keywords Ant colony optimization · Artificial immune system · Bioinspired protocol · Mobile ad hoc networks · Routing.

1 Introduction

Optimization problems can be solved by artificial immune systems. These problems we face with these kinds of problems daily: the efficiency improvement of the resources of the devices, find the shortest path between two points, distribute the resources in the system uniformly.

D. Rupérez Cañas · A. L. Sandoval Orozco · L. J. García Villalba (✉)
Department of Software Engineering and Artificial Intelligence (DISIA), Group of Analysis Security and Systems (GASS), School of Computer Science, Office 431 Universidad Complutense de Madrid (UCM) Calle Profesor José García Santesmases s/n Ciudad Universitaria, 28040 Madrid, Spain
e-mail: javiergv@fdi.ucm.es

D. Rupérez Cañas
e-mail: delfinrc@fdi.ucm.es

A. L. Sandoval Orozco
e-mail: asandoval@fdi.ucm.es

One of the optimization algorithms based on the colony of ants [1] and that relies on the intelligence swarm [2] has been frequently cited in the literature. It is inspired by the behavior of ants at the time of obtaining the food, and in many areas, it is applied.

ACO algorithms are composed by agents that work without the need of a centralized control structure, in such a way that the interactions local to each agent and its neighbors allow them to communicate in an autonomous way. These algorithms can be used to resolve routing problems, being suitable for highly dynamic environments. We present improvements in the optimization of protocol AntOR [3], in its disjoint-link version, and we show its relationship with the artificial immune systems. A work related to immune systems is [4]. In this work, the authors try to solve problems of misbehaviors in mobile ad hoc networks (MANETs) taking into account the artificial immune systems, but they have used the standard protocol DSR which is reactive and does not exploit the properties of the hybrids.

We structure the rest of article as follows: In Sect. 2, we explain our proposal as a view point of immunology. In Sect. 3, the simulation results in a dynamic environment are exposed, comparing them with the standard protocol AODV. Section 4 presents conclusions.

2 Proposed Algorithm

We present the hybrid (mix between reactive and proactive parts) routing protocol AntOR [3] with the following characteristics:

- Disjoint-link and disjoint-node protocol [5].
- Separation between the pheromones values in the diffusion process.
- Use of the distance metric in the proactive path exploration.

AntOR provides two versions in its design: the disjoint-link (AntOR-DLR), in which the links are not shared, and disjoint-node (AntOR-DNR), in which the nodes are not shared. Every disjoint-node is also a disjoint-link, but not vice versa. Both types of disjoint routes have the following advantages:

1. A failure in one node only affects a path, not the entire network.
2. Load balancing is better because there are not repeated routes on the disjoint property.

However, the use of such routes needs more resources by not sharing the links or nodes.

This algorithm is applied to mobile ad hoc networks (MANETs), and the optimization might be addressed by ideas of immunology. This algorithm is modeled in the following way:

- **Body:** The entire mobile ad hoc network.
- **Antibody:** Address pairs consisting of the “next hop” and “destination.”

- **Antigen:** Destination of the data packet.
- **Matching:** Correspondence between the associated destination with the data packet and destination field of a pair which belongs to an antibody.
- **Affinity:** Heuristic value (regular pheromone).

The new technique used and reflected in the simulation results is reduction in overload of the system through proactive agents that do not need virtual pheromone routes. These agents create alternative routes and go from neighbor to neighbor until reaching the destination node. At the time of selecting the next hop, they take into account the maximum value of regular pheromone to such a one-hop neighbor. Alternative routes are achieved with this technique up to a limit which are selected previously and which are disjoint because those routes do not belong to the main route.

Another technique is related to fault tolerance. When a fault in a control message (an agent of our algorithm) is detected, we trigger a mechanism of neutralization process. This makes that in highly dynamic environments, we have to trigger more neutralization mechanisms by sending agents to repair the route or notify to the precursors of the node that detects the failure until reaching the source of the data session, indicating that the route is disconnected. This implies an overhead in packets and bytes. To fix this, we use a new technique that checks if exists route (regular pheromone value greater than zero) to the neighbor whose we want to transmit, seeing this information in the routing table. If path exists, we send the control message; otherwise, the agent does not send. Thus, this prevents the failure neutralization, and it reduces overhead.

3 Simulation Results

Performance metrics are delivered data packet ratio and overhead in packets.

The characteristics of the simulations in network simulator NS-3 were as follows: We used 100 nodes randomly distributed and configured with a transmission range of 300 m. The nodes are moved according to the *Random Way Point* (RWP) pattern, varying pause time from a low of 0 to 240 s at intervals of 60 s. The scenario was rectangular with dimensions 3000, 1000 m. The speed was variable from a minimum of 0 to 8 m/s. It used 10 random data sessions using the application protocol *Constant Bit Rate* (CBR) beginning to send data at random from 0 s to a maximum of 60 s. The sending rate was 512 bit/s, that is, sending a packet of 64 bytes per second. The maximum simulation time was established to 300 s. It employed a total of 10 runs in the experiment.

Figure 1 shows how the data packet ratio in our proposed protocol AntOR-DLR is better than that in AODV at all time. The ratio is an important metric of effectiveness.

Figure 2 shows how the overhead in AntOR is practically the same as AODV.

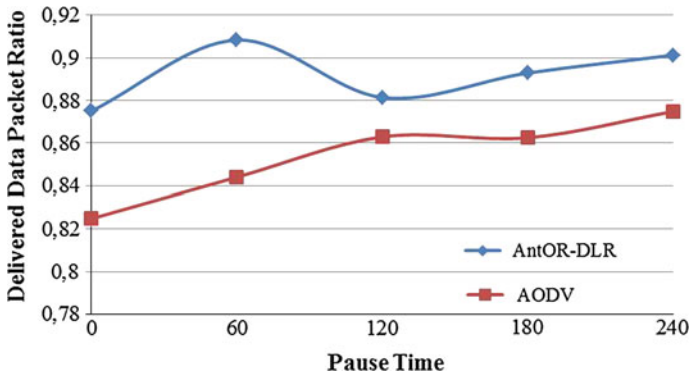


Fig. 1 Pause time versus ratio

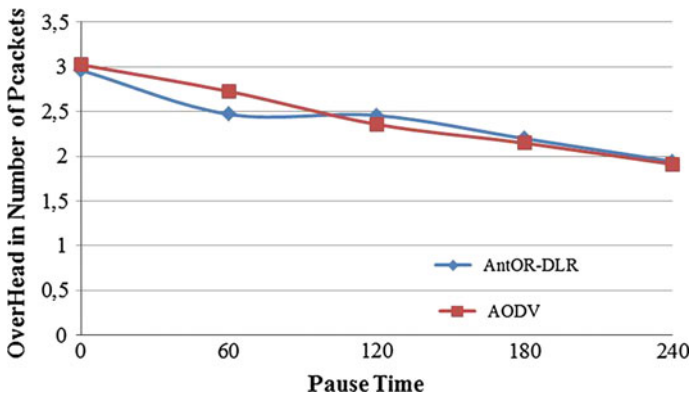


Fig. 2 Pause time versus overhead in packets

4 Conclusion

In this article, we have presented an ACO bioinspired algorithm and provided new optimization techniques, relating it with concepts of artificial immune systems. Finally, we can see that these new techniques reduction in the overhead in the system and fault tolerance originate this version AntOR-DLR to behave better than AODV according to metrics of ratio and overhead in number of packets.

Acknowledgments This work was supported by the Ministerio de Industria, Turismo y Comercio (MITyC, Spain) through the Project Avanza Competitividad I+D+I TSI-020100-2011-165, and the Agencia Española de Cooperación Internacional para el Desarrollo (AECID, Spain) through Acción Integrada MAEC-AECID MEDI-TERRÁNEO A1/037528/11.

References

1. Dorigo M (1992) Optimization. Learning Nat Algorithms Doctoral Thesis. Politecnico di Milano, Italie
2. Kennedy J (2001) Swarm intelligence. Morgan Kaufmann Publishers
3. García Villalba LJ, Rupérez Cañas D, Sandoval Orozco AL (2010) Bioinspired routing protocol for mobile ad hoc networks. *IET Commun* 4(18):2187–2195
4. Le Boudec L, Sarajanović S (2004) An artificial immune system approach to misbehavior detection in mobile Ad-Hoc networks. In: Proceedings of the first international workshop on biologically inspired approaches to advanced information technology (Bio-ADIT 2004), Lausanne, Switzerland, January 29–30, pp 96–111
5. Rupérez Cañas D, Sandoval Orozco AL, García LJ, Kim T-H (2011) A comparison study between AntOR-Disjoint node routing and AntOR-Disjoint link routing for mobile Ad Hoc networks. *Commun Computer Inf Sci (CCIS)* 263:300–304

Bayesian Regularization BP Neural Network Model for the Stock Price Prediction

Qi Sun, Wen-Gang Che and Hong-Liang Wang

Abstract It is an important research issue to improve the generalization ability of the neural network in the research of artificial neural network. This paper proposes the Bayesian regularization method to optimize the training process of the back propagation (BP) neural network, so that the optimized BP neural network model can predict new data in the BP neural network to a larger extent. Based on the experiments in which Bayesian regularization BP neural network is employed to predict the stock price series, and through the establishment of the stock customer transaction model network structure, an experimental program is selected to make an empirical analysis of the closing price data of Shanghai Stock in 800 trading days, the results of which show that the Bayesian regularization method has a better generalization ability.

Keywords Outlier · Neural networks · Bayesian regularization · The stock price · The generalization ability

Q. Sun (✉) · W.-G. Che

Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, People's Republic of China
e-mail: 531506873@qq.com

W.-G. Che

e-mail: wgche@yahoo.com

H.-L. Wang

Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang 110168, People's Republic of China
e-mail: wanghl@sict.ac.cn

1 Introduction

Network performance is mainly measured by its generalization ability. Generalization is defined as the extent to which the model trained in the training set can predict the correct output of the new examples.

In 1986, Rumelhart et al. proposed a back propagation (back propagation, BP) algorithm, which propagates the error of the output layer from back to front layer by layer. BP neural network essentially realizes the mapping from input to the output, and the mathematical theory has proved that it has the ability to realize any complex nonlinear mapping, which makes it particularly suitable for solving the complex problems of the internal mechanism. BP network is able to learn from the samples with correct answers and extract a “reasonable” solution automatically, that is, it has the ability of self-learning. Besides, it has the capability to conduct promotion and generalization. However, there also exists the contradiction between the prediction ability of the BP neural network (also called generalization ability or promotion ability) and its training ability (also called approximation ability or learning ability). Generally speaking, when the training ability is poor, the prediction ability is also not very good. To a certain extent, the predictive ability is enhanced with the improvement of training capabilities. However, this trend has a limit. When this limit is reached, with the improvement of the training ability, the predictive ability will decrease, which is the so-called over-fitting phenomenon [1]. In this case, the network learns too many details of the sample and cannot reflect the internal law of the sample. Considering these shortcomings, this paper adopts Bayesian regularization BP neural network forecasting model and conducts a data experiment based on the stock price.

The stock market is a barometer of the national economy. When the stock market investors want to make an analysis or a decision, they need a lot of data, based on which they can explore the operating rules and the future trend. The stock price series can be seen as a time series with white noise in the nonlinear system [2]. In this paper, L-M optimization neural network [3] and Bayesian regularization neural network are selected to make a comparative experiment in MATLAB. Based on the data, this paper attempts to prove that under the same conditions, the results show that Bayesian regularization of BP neural network can predict the stock index more effectively. Compared with other improved algorithms, it shows better generalization ability.

2 BP Neural Network

BP algorithm is divided into two stages: The first stage is a forward process, in which the output values of each unit are calculated layer by layer; the second stage is a reverse process, in which the error of each hidden layer is calculated, based on which the weight of the previous layer is corrected. This process solves the problem that the multilayer perceptron cannot calculate the error of the hidden layer.

2.1 Definition

If the error back propagation signal increased on the basis of the multilayer perceptions, then the nonlinear information can be processed, and such a network is called the forward network of the error back propagation (BP) [4]. The whole process can be summarized as “mode forward propagation → error back propagation → Memory Training → learning convergence.”

2.2 Performance Index

The index performance of the BP neural network is the quantitative criterion, by which the mean square error in response to the network is used to measure its performance. That is,

$$E_d = \frac{1}{n} \sum_{p=1}^n (t_p - a_p)^2 \quad (1)$$

where E_d is the mean square error, n denotes the total number of samples, t_p is the desired output value of the p th training group, a_p the actual output value of the p th training group.

3 Bayesian Regularization BP Neural Network

3.1 Regularization

Assume C to be an irreducible plane algebraic curve, S is a collection of C 's singularity. If there is a compact Riemann surface and holomorphic mapping making the following conditions satisfied:

$$\delta(C^*) = C;$$

$$\delta^{-1}(S) \text{ is a finite set of points;}$$

$$\delta : \frac{C^*}{\delta^{-1}(S)} \rightarrow \frac{C}{S} \text{ is a one-to-one mapping.}$$

then (C^*, δ) will be claimed as the regularization of C [5].

As a matter of fact, in the regularization approach, the curve branches with different tangents are separated at the point of singularity of the irreducible plane algebraic curve, so that the singularity can be eliminated. This paper adopts the regularization method to modify the training performance function of the neural network in order to improve its generalization ability. Set neural network training samples as $D = (x_i, t_i)$, $i = 1, 2, \dots, n$, where n is the total number of the samples; S is the network structure; W is the parameter vector; $f(\bullet)$ is the actual output

of the network; k is the network output; m is the total number of the parameters. The network error function is as follows:

$$E_D = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^k [f(x_i, W, S) - t_i]^2 \quad (2)$$

$$\text{Weight attenuation function : } E_W = \frac{1}{2} \|W\|^2 = \frac{1}{2} \sum_{i=1}^m w_i^2 \quad (3)$$

The overall error function is $F(W) = \alpha E_W + \beta E_D$, where the parameters α and β control the distribution pattern of the weight and the threshold value. If $\alpha \ll \beta$, then emphasis will be laid on the reduction in the training error and there may appear the phenomenon of “over-fitting”; if $\alpha \gg \beta$, then the emphasis will be laid on the restriction of the weight value of the network, and network scale will shrink automatically, and the error may become larger, leading to the phenomenon of “under-fitting.” Thus, it is necessary to make a compromise in the selection of the parameters α and β . During the network training, in order to reduce the complexity of the network structure and the error, it is necessary to minimize the objective function. In this paper, the authors select the method of Bayesian regularization to choose the parameters α and β .

3.2 Bayesian Regularization

The Bayesian regularization BP neural network can automatically adjust the size of α and β in the training process, until they reach the optimization. Assume that the noise and the weight vectors both follow the Gaussian distribution, then

$$p(D|w, \beta, S) = \frac{\exp(-\beta E_d)}{Z_n(\beta)}, \quad p(w|\alpha, S) = \frac{\exp(-\alpha E_w)}{Z_m(\alpha)} \quad (4)$$

wherein, $Z_n(\beta) = \left(\frac{\pi}{\beta}\right)^{\frac{n}{2}}$, $Z_m(\alpha) = \left(\frac{\pi}{\alpha}\right)^{\frac{m}{2}}$.

Set network weight w as a random variable. The posterior probability density of the weights after learning set through the Bayesian equation [6]:

$$p(w|D, \alpha, \beta, S) = \frac{p(D|w, \beta, S)p(w|\alpha, S)}{p(D|\alpha, \beta, S)} \quad (5)$$

wherein $p(w|\alpha, S)$ is the priori probability density function of the weight vectors. $p(D|w, \beta, S)$ is the output probability density function when the weights are given, and $p(D|\alpha, \beta, S)$ is the normalization factor. Since the standardized factor is unrelated to the weight vector w , then

$$p(w|D, \alpha, \beta, S) = \frac{e^{-(\alpha E_w + \beta E_d)} p(D|w, \beta, S) p(w|\alpha, S)}{p(D|\alpha, \beta, S)} = \frac{e^{-F(w)}}{Z(\alpha, \beta)}. \quad (6)$$

where $Z(\alpha, \beta) = \frac{p(D|\alpha, \beta, S)}{Z_n(\beta)Z_m(\alpha)}$, optimal weight vector has the maximum posteriori probability. When $Z(\alpha, \beta)$ is determined, the maximum posteriori probability will be equivalent to $F(w)$. When the minimum value is gotten from $F(w)$, the corresponding weight value (including the threshold value) is w^* , the optimal solutions of α and β can be obtained: $\alpha = \frac{r}{2E_w(w^*)}$, $\beta = \frac{m-r}{2E_d(w^*)}$ where r is the effective number of the network parameters determined by the trained learning set.

The employment of Bayesian regularization method is, in essence, the use of Bayesian mathematical statistics method to automatically determine the regularization parameters.

3.3 The Algorithm of Bayesian Regularization BP Neural Network

The following is a list of five steps:

- Step 1: Determine the network structure. Set the parameters α and β and initialize them. Generally, set $\alpha = 0$ and $\beta = 1$;
- Step 2: Training Network. “tansig” is adopted in the design of the network input function, “purelin” in the output function, and “trainbr” in the training function. The input vector and the destination vector on the training data are normalized using the linear conversion function, and the equation is $y_{in} = \frac{x_{in} - \text{min value}}{\text{max value} - \text{min value}}$ where x and y are the values before and after the conversion, respectively. The normalized output equation is $y_{out} = x_{out}(\text{max value} - \text{min value}) + \text{min value}$;
- Step 3: Find the minimum point. The Hessian array is obtained using the Gauss–Newton approximation method: $\nabla^2 F(w^*) \approx 2\beta J^T J + 2\alpha I_m$, and the number of the effective parameters r is calculated, where J is the Jacobian matrix of E_D at the point w^* ;
- Step 4: Judge the objective function. Through $\alpha = \frac{r}{2E_w(w^*)}$, $\beta = \frac{m-r}{2E_d(w^*)}$, the new estimated values of α and β are calculated.
- Step 5: End the training. Repeat the steps from 1 to 4, until the desired accuracy is reached.

4 Experimental Analysis

In this experiment, stock No. 900908 of Shanghai Stock Exchange is taken as the example. And 800 closing prices from May 8, 2009, to August 28, 2012, are selected as the experimental data. And L-M optimization BP neural network

algorithm and Bayesian regularization BP neural network are adopted to predict the closing price of the stock. Based on the sequence set containing weight, which is generated through relevance principle and sequential mode, a customer transaction behavior model is then generated [7]:

$$BH = \begin{pmatrix} S_{11} & \cdots & S_{1a} & 0 & 0 & 0 \\ S_{21} & \cdots & \cdots & \cdots & \cdots & S_{2n} \\ S_{31} & \cdots & \cdots & S_{3c} & 0 & 0 \\ \vdots & \cdots & \cdots & \vdots & \cdots & \vdots \\ S_{k1} & \cdots & \cdots & S_{kp} & \cdots & 0 \end{pmatrix}$$

BH is a $k \times n$ matrix, where each row represents a piece of sequence mode, which reveals the contextual sequence between different stocks traded by investors. From the first row to the last row, the supporting degree of the sequence decreases progressively. Each element in the matrix can be a stock or the set of several stocks. For an investor, the trading security products in the stock swap constitute a sequence, that is, s_1, s_2, \dots, s_n . In this sequence, due to the factors such as the positions status, personal preferences, the status of s_n may be associated with its formers such as the status of $s_{n-1}, s_{n-2}, s_{n-3}$. However, if we examine the statistical distribution of the overall data warehouse from the perspective of the investors as a whole, at the time of t , the investors trading security varieties status is s_n . At the moment of $t-1$, the investors trading securities varieties of state are s_{n-1} . As is indicated in the data exploration structure of the sequence mode, the trading security varieties status s_n just related to the status at the time of $t-1$ (Table 1).

These data suggest that s_n and s_n show a high degree of correlation (96.33 %) and but its correlation with s_{n-1}, s_{n-2} is rather weak, which are 3.38 and 0.29 %, respectively. From a statistical perspective, the customer transaction behavior mode is a typical Markov chain [8]. Assume that the stock price change mainly due to the customer transaction behavior. Then, we choose the closing price of predicted two days before to predict the closing price of the third day.

The 500 times of L-M optimization algorithm training process are shown in Figs. 1, 2, and 3:

Figure 1 shows the network performance optimized by L-M optimization algorithm, and it gets the best training performance after 114 iterations.

Figure 2 shows that the gradient is the gradient of the error surface, when the gradient reached a value can be the end of the training. The variable mu determines how learning is done, based on Newton’s method or gradient method. As long as the error increases by iteration, mu will increase, until the error does not increase; however, if mu is too large, it will make the learning stop that occurs when the minimum error has been found, and it is why when mu reached the maximum, it must stop learning. Validation checks indicate the effect of L-M optimization network evolution.

Bayesian regularization training process is shown in Figs. 4, 5, and 6:

Table 1 Data analysis is conducted for the exploration process, the minimum supporting degree of which is set to 5 %, and the data are as follows

Sequence mode	2	3	More than 3	Sum
Number	59,87	210	18	62,15
Percentage (%)	96.33	3.38	0.29	100

Fig. 1 Mean square error (MSE)

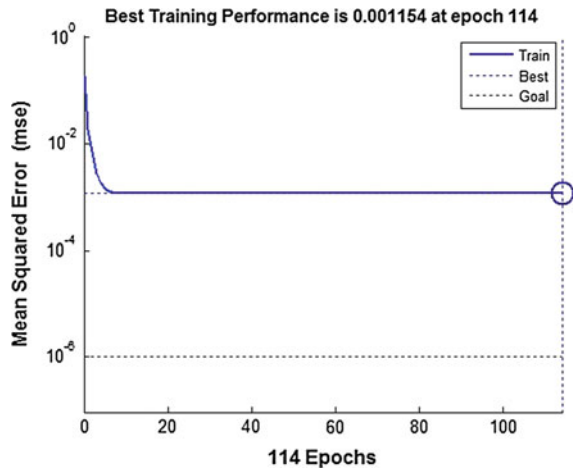


Fig. 2 Training process diagram

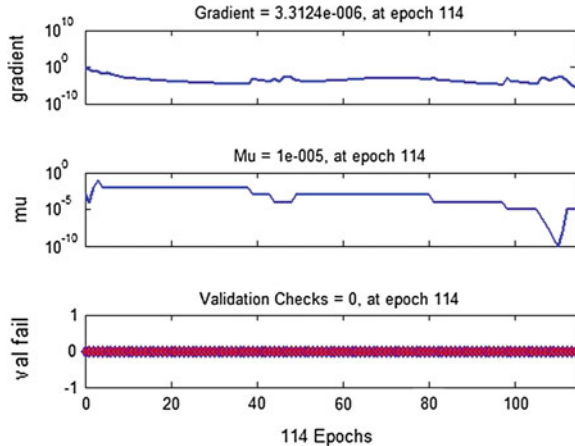


Figure 4 shows the network performance optimized by Bayesian regularization algorithm, and it gets the best training performance after 145 iterations.

Figure 5 shows four kinds of parameters about the Bayesian regularization training. The sense of squared keeps two sets of data between the true and the estimated value be compared with each other after a positive number, only to get the gap but not to get the specific relationship between them.

The comparison of the errors between the two algorithms is shown in Fig. 7:

Fig. 3 Correlation diagram

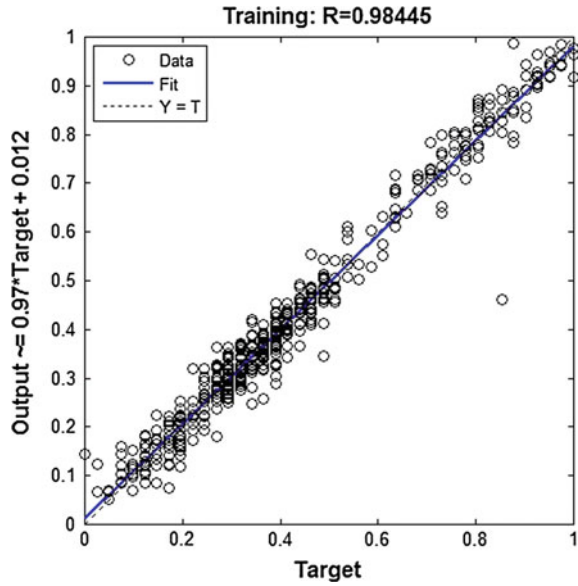


Fig. 4 Sum-squared error (SSE)

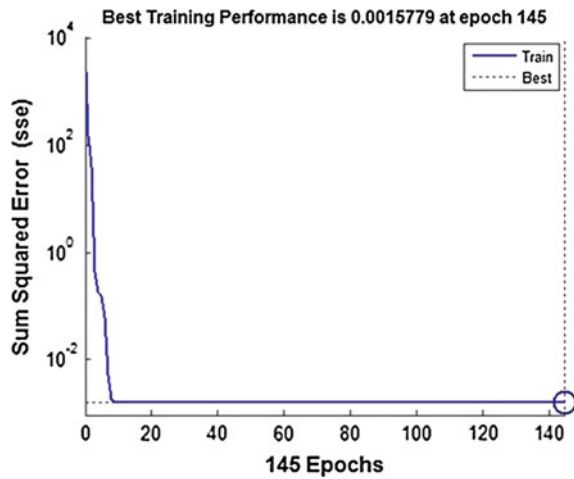


Figure 7 displays that the error fluctuations of L-M optimization algorithm are more dramatic than those of BP algorithm with Bayesian regularization.

Figures 1 and 4 show that convergence speed of the L-M algorithm is higher than that of the regularization Bayesian algorithm. But Figs. 3 and 6 indicate that the fitting curve of the L-M algorithm is not smooth. Although the over-fitting network has less training residual, there are many discrete points, and the structure is bloated. Thus, enough information cannot be obtained, leading to a useless prediction of many unknown data, poor generalization ability, as well as poor applicability. Figure 7 shows that the prediction effect of the L-M algorithm is

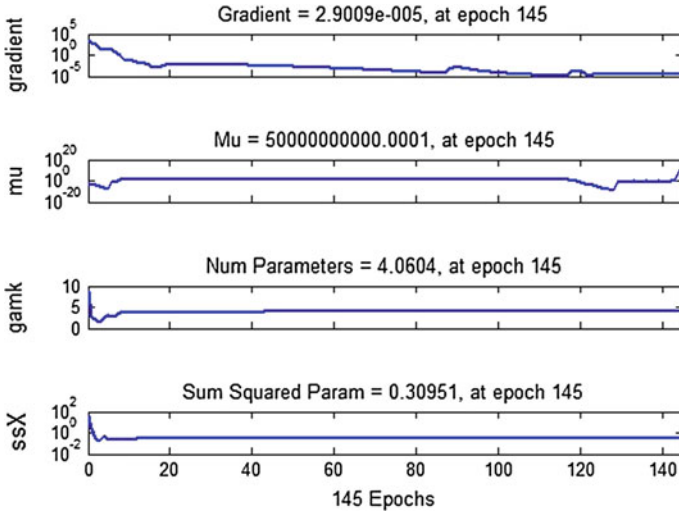
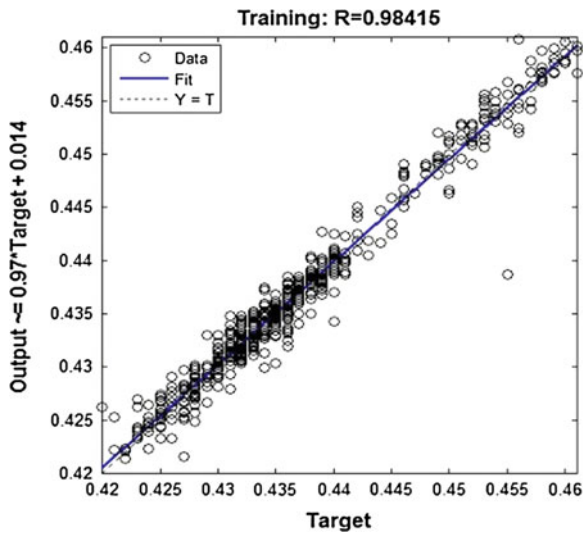


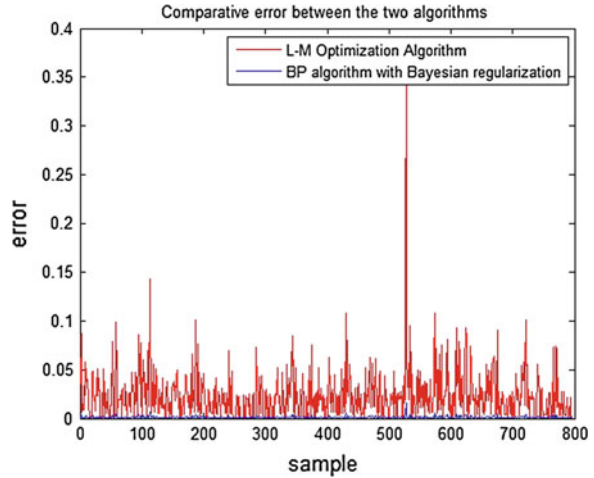
Fig. 5 Training process

Fig. 6 Correlation diagram



unstable. But the BP network after the Bayesian regularization has a stable prediction ability, and the prediction experiment of the closing price of all stocks indicates that the BP neural network after the Bayesian regularization presents more accurate prediction of the B-share of the Shanghai Stock Exchange, and greater error in the prediction of the A shares but with a more obvious generalization ability.

Fig. 7 Comparative error



5 Conclusions

In this paper, a Bayesian regularization BP neural network model is introduced to improve the generalization ability of the neural network. The stock data are used in prediction, and a good effect is reached. And through the comparison of the experimental diagrams of the traditional prediction method and the Bayesian regularization BP neural network model, it is proved that the generalization ability is enhanced. For general algorithm, the mean square error function is the objective function. The issue of weights cannot be optimized. As for the Bayesian regularization BP neural network model, the weight value is added in the objective function, by way of which the parameters are adjusted automatically and the network structure is optimized. Thus, it shows better predictability, finally improving the generalization capability of the network.

References

1. Jin L, Kuang X, Huang H, Qin Z, Wang Y (2005) Study on the overfitting of the artificial neural network forecasting model. *Acta Meteorol Sinica* 19(2):216–225
2. Wang B, Zhang F (2005) Comparison of artificial neural network and time series model for forecasting stock prices. *J Wuhan Automot Polytech Univ* 27(6):69–73
3. Zhang B, Yuan S, Cheng L, Yuan J, Cong X (2004) Model for predicting crop water requirements by using L-M optimization algorithm BP neural network. *Trans Chin Soc Agric Eng* 20(6):73–76
4. Choudhary A, Rishi R (2011) Improving the character recognition efficiency of feed forward BP neural network. *Int J Comput Sci Inf Technol (IJCSIT)* 3(1)
5. Lü S (2011) *Statistical learning algorithms for regression and regularized spectral clustering*. University of Science and Technology of China, Hefei
6. Wu G, Tao Q, Wang J (2005) Support vector machines based on posteriori probability. *J Comput Res Dev* 42(2):196–202

7. Feng W, Pengfei S (2000) Client-transaction-behavior analysis using conceptual clustering. *Microcomput Appl* 16(5):107–110
8. Sun B, Li T, Wang B (2011) Neural network forecasting model based on stock market sensitivity analysis. *Comput Eng Appl* 47(1):26–31

Adaptive Fuzzy Control for Wheeled Mobile Manipulators with Switching Joints

Zhijun Li

Abstract The switching joints can be switched to either active (actuated) or passive (under-actuated) mode as needed [1]; in this paper, dynamic coupling switching control incorporating fuzzy logic systems is developed for wheeled mobile manipulators with switching joints. The switching actuated robot dynamics is an mixed under-actuated and actuated model. The fuzzy logic systems are employed to approximate the high dimension unmodelled dynamics. Considering the joint switch as an event, the event-driven switching control strategy is used to ensure that the system outputs track the given bounded reference signals within a small neighborhood of zero.

1 Introduction

The switching joint shown in Fig. 1 was first proposed in Li et al. [1], which is equipped with one clutch, when the clutch is released, the link is free, and the passive link is directly controlled by the dynamic coupling of mobile manipulators, and when it is on, the joint is actuated by the motor. The robot with switching joints is called the switching actuated robot.

Switching actuated mobile manipulator is the robot manipulators consisting of switching joints mounting on a wheeled mobile robot. These systems are intrinsically nonlinear, and their dynamics will be described by nonlinear differential equations. The switching joints in the free mode, which can rotate freely, can be indirectly driven by the effect of the dynamic coupling between the active and passive joints. The zero torque at the switching joints results in a second-order nonholonomic constraint.

Z. Li (✉)

College of Automation Science and Engineering, South China University of Technology, Guangzhou, China
e-mail: zjli@ieee.org

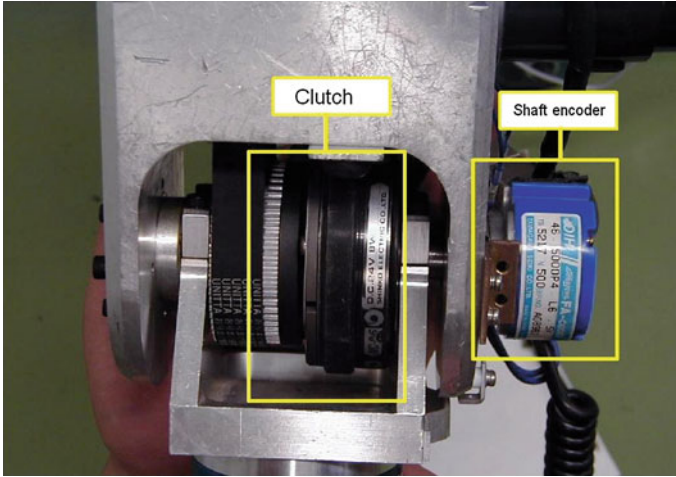


Fig. 1 The switching joint

Switching control system has recently received much attention due to its applicability to many physical systems exhibiting switching in nature. The switching joint is a typical switching system [2], which has not been investigated until now. In the switching control for switching joints, the control can be switched among several controllers, and each controller is designed for a specific nominal mode of the switching joint.

Recently, fuzzy logic control has found extensive applications for complex and ill-defined systems, especially in the presence of incomplete knowledge of the plant or the situation where precise control action is unavailable. Based on the universal approximation theorem [3], many stable fuzzy adaptive control schemes have been developed for unknown single-input and single-output (SISO) nonlinear systems [4], and multiple-input and multiple-output (MIMO) nonlinear systems [5], and achieve stable performance criteria.

Therefore, in this paper, we consider switching control incorporating fuzzy logic for switching actuated mobile manipulators in the presence of parametric and functional uncertainties in the dynamics. We propose a state-dependent switching controller for motion control of mobile manipulator with joint switch. The switching signal is assumed to be generated by an event, which leads to two different switching controls. In each stage, the system reaches a subspace of the whole space and converges to the desired trajectory. We prove the convergence of the closed-loop system to the equilibrium point for every stage.

2 System Description

2.1 The Dynamics of Wheeled Mobile Manipulators

The dynamics of an n DOF mobile manipulator mounted on a two-wheeled driven mobile platform can be expressed as the following:

$$D(q)\ddot{q} + C(q, \dot{q})\dot{q} + G(q) + d(t) = B(q)\tau + f \tag{1}$$

where

$$D(q) = \begin{bmatrix} D_v & D_{va} & D_{vh} \\ D_{av} & D_a & D_{ah} \\ D_{hv} & D_{ha} & D_h \end{bmatrix}, \quad C(q, \dot{q}) = \begin{bmatrix} C_v & C_{va} & C_{vh} \\ C_{av} & C_a & C_{ah} \\ C_{hv} & C_{ha} & C_h \end{bmatrix}, \quad G(q) = \begin{bmatrix} G_v \\ G_a \\ G_h \end{bmatrix}$$

$$d(t) = \begin{bmatrix} d_v \\ d_a \\ d_h \end{bmatrix}, \quad B(q)\tau = \begin{bmatrix} \tau_v \\ \tau_a \\ \tau_h \end{bmatrix}, \quad f = \begin{bmatrix} J_v^T \lambda_n \\ 0 \\ 0 \end{bmatrix},$$

q is the generalized coordinates for the mobile manipulators with $q = [q_v^T, q_a^T, q_h^T]^T \in R^n$; q_v is the generalized coordinates for the mobile platform with $q_v = [x, y, \theta]^T \in R^{n_v}$; x, y are the position coordinates, and θ is the heading angle; q_a are the coordinates of the active joints with $q_a \in R^{n_a}$; q_h denote the coordinates of the switching joints with $q_h \in R^{n_h}$; $D(q)$ is the symmetric inertia matrix; D_v, D_a, D_h are the inertia matrices for the mobile platform, the active links, and the switching links, respectively; $C(q, \dot{q})$ are the centripetal and Coriolis torques; C_v, C_a, C_h are the centripetal and Coriolis torques for the mobile platform, the active links, and the switching links, respectively; $G(q)$ is the gravitational torque vector with $G(q) = [G_v, G_a, G_h]^T \in R^n$; $d(t)$ are the external disturbances with $d(t) = [d_v, d_a, d_h] \in R^n$; d_v, d_a, d_h are the external disturbances on the mobile platform, the active links, and the switching links, respectively; $B(q)$ is the known full rank input transformation matrix with $B(q) \in R^{n \times m}$; τ are the control inputs with $\tau = [\tau_v, \tau_a, \tau_h] \in R^m$; f is the external force on the mobile manipulators with $f \in R^l$.

2.2 Reduced System

The vehicle subjected to nonholonomic constraints can be expressed as

$$J_v \dot{q}_v = 0 \tag{2}$$

The effect of the constraints can be viewed as a restriction of the dynamics on the manifold Ω_n as $\Omega_n = \{(q_v, \dot{q}_v) | J_v \dot{q}_v = 0\}$.

Constraints (2) imply the existence of vector $\dot{\eta} \in R^{n_v-l}$, such that

$$\dot{q}_v = S(q_v)\dot{\eta} \quad (3)$$

Considering (3) and its derivative, the dynamics of mobile manipulator can be expressed as

$$\mathcal{D}(\zeta)\ddot{\zeta} + \mathcal{C}(\zeta, \dot{\zeta})\dot{\zeta} + \mathcal{G}(\zeta) + d(t) = \mathcal{U} \quad (4)$$

where

$$\begin{aligned} \mathcal{D}(\zeta) &= \begin{bmatrix} S^T D_v S & S^T D_{va} & S^T D_{vh} \\ D_{av} S & D_a & D_{ah} \\ D_{hv} S & D_{ha} & D_h \end{bmatrix}, \quad \zeta = \begin{bmatrix} \eta \\ q_a \\ q_h \end{bmatrix}, \quad \mathcal{G}(\zeta) = \begin{bmatrix} S^T G_v \\ G_a \\ G_h \end{bmatrix}, \\ d_1(t) &= \begin{bmatrix} H^T d_v \\ d_a \\ d_h \end{bmatrix}, \quad \mathcal{C}(\zeta, \dot{\zeta}) = \begin{bmatrix} S^T D_v \dot{S} + S^T C_v S & S^T C_{va} & S^T C_{vh} \\ D_{av} \dot{S} + C_{av} S & C_a & C_{ah} \\ D_{hv} \dot{S} + C_{hv} S & C_{ha} & C_h \end{bmatrix}, \\ \mathcal{U} &= \begin{bmatrix} S^T \tau_v \\ \tau_a \\ \tau_h \end{bmatrix}. \end{aligned}$$

Remark 1 In this paper, we choose $\dot{\zeta} = [v, \omega, \dot{q}_a^T, \dot{q}_h^T]^T$, and $\eta = [v, \omega]^T$, where v is the forward velocity of the mobile platform; and ω is the rotation velocity of the mobile platform.

Assumption 2.1 For simplicity, the disturbance d_1 is bounded.

2.3 Physical Properties for Switching Modes

Actuated Switching Joints Assume that the switching joint $q_h = \zeta_3$ is known, we select the group of the actuated joints and $\zeta_1 = q_a$ from all actuated joints including the mobile platform such that the dimension of ζ_1 and ζ_3 are equal, and the remaining actuated variables are grouped as ζ_2 . Partition (4) in quantities related to the active joints, the switching joints, and the remaining joints of the mobile manipulators as

$$\begin{aligned} \mathcal{D}(\zeta) &= \begin{bmatrix} D_{11} & D_{12} & D_{13} \\ D_{21} & D_{22} & D_{23} \\ D_{31} & D_{32} & D_{33} \end{bmatrix}, \quad \mathcal{C}(\zeta, \dot{\zeta})\dot{\zeta} = \begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix} = \begin{bmatrix} C_{11}\dot{\zeta}_1 + C_{12}\dot{\zeta}_2 + C_{13}\dot{\zeta}_3 \\ C_{21}\dot{\zeta}_1 + C_{22}\dot{\zeta}_2 + C_{23}\dot{\zeta}_3 \\ C_{31}\dot{\zeta}_1 + C_{32}\dot{\zeta}_2 + C_{33}\dot{\zeta}_3 \end{bmatrix}, \\ \mathcal{G} &= \begin{bmatrix} G_1 \\ G_2 \\ G_3 \end{bmatrix}, \quad \mathcal{P} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix}, \quad \mathcal{U} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}. \end{aligned}$$

Considering the property of the above mechanical system, we list the following properties [7] for the active switching joints:

Property 1 The inertia matrix $\mathcal{D}(\zeta)$ is symmetric, positive definite, and bounded.

Property 2 The matrix $\dot{\mathcal{D}} - 2\mathcal{C}$ is skew-symmetric.

Unactuated Switching Joints If switching joint is under-actuated, then $u_3 = 0$, it needs to assume that $n_v + n_a > n_h$ such that the switching links can be controlled. If $n_h = 1$, that means, there exists one switching joint, even if $n_a = 0$, and since $n_v = 2$, the system still can be controlled. In order to make ζ_3 controllable, we assume that matrices D_{13} and D_{31} are of full rank and it is obvious that D_{11}^{-1} exists. However, if D_{13} and D_{31} are not of full rank, while D_{23} and D_{32} have full rank, which means that ζ_3 will be coupled with ζ_2 , we only need to exchange ζ_1 with the vector ζ_2 . In this paper, for simplification, we assume $D_{13} = D_{31}$ being of full rank. After some simple manipulations, we can obtain three dynamics as

$$D_{11}\ddot{\zeta}_1 = u_1 - C_1 - G_1 - d_{11} - D_{12}\ddot{\zeta}_2 - D_{13}\ddot{\zeta}_3 \tag{5}$$

$$\begin{aligned} (D_{22} - D_{21}D_{11}^{-1}D_{12})\ddot{\zeta}_2 + (D_{23} - D_{21}D_{11}^{-1}D_{13})\ddot{\zeta}_3 \\ + C_2 + G_2 + d_{12} - D_{21}D_{11}^{-1}C_1 - D_{21}D_{11}^{-1}G_1 \\ - D_{21}D_{11}^{-1}d_{11} = u_2 - D_{21}D_{11}^{-1}u_1 \end{aligned} \tag{6}$$

$$\begin{aligned} (D_{32} - D_{31}D_{11}^{-1}D_{12})\ddot{\zeta}_2 + (D_{33} - D_{31}D_{11}^{-1}D_{13})\ddot{\zeta}_3 \\ + C_3 + G_3 + d_{13} - D_{31}D_{11}^{-1}C_1 - D_{31}D_{11}^{-1}G_1 \\ - D_{31}D_{11}^{-1}d_{11} = -D_{31}D_{11}^{-1}u_1 \end{aligned} \tag{7}$$

Let $\mathcal{A} = D_{22} - D_{21}D_{11}^{-1}D_{12}$, $\mathcal{B} = D_{23} - D_{21}D_{11}^{-1}D_{13}$, $\mathcal{J} = D_{32} - D_{31}D_{11}^{-1}D_{12}$, $\mathcal{L} = D_{33} - D_{31}D_{11}^{-1}D_{13}$, $\mathcal{E} = (C_{22} - D_{21}D_{11}^{-1}C_{12})\dot{\zeta}_2 + (C_{23} - D_{21}D_{11}^{-1}C_{13})\dot{\zeta}_3$, $\mathcal{F} = (C_{32} - D_{31}D_{11}^{-1}C_{12})\dot{\zeta}_2 + (C_{33} - D_{31}D_{11}^{-1}C_{13})\dot{\zeta}_3$, $\mathcal{H} = (C_{21} - D_{21}D_{11}^{-1}C_{11})\dot{\zeta}_1 + G_2 + d_{12} - D_{21}D_{11}^{-1}G_1 - D_{21}D_{11}^{-1}d_{11}$, $\mathcal{K} = (C_{31} - D_{31}D_{11}^{-1}C_{11})\dot{\zeta}_1 + G_3 + d_{13} - D_{31}D_{11}^{-1}G_1 - D_{31}D_{11}^{-1}d_{11}$. Then, we can rewrite (5), (6), and (7) as

$$D_{11}\ddot{\zeta}_1 = u_1 - C_1 - G_1 - d_{11} - D_{12}\ddot{\zeta}_2 - D_{13}\ddot{\zeta}_3 \tag{8}$$

$$\mathcal{A}\ddot{\zeta}_2 + \mathcal{B}\ddot{\zeta}_3 + \mathcal{E} + \mathcal{H} = -D_{21}D_{11}^{-1}u_1 + u_2 \tag{9}$$

$$\mathcal{J}\ddot{\zeta}_2 + \mathcal{L}\ddot{\zeta}_3 + \mathcal{F} + \mathcal{K} = -D_{31}D_{11}^{-1}u_1 \tag{10}$$

Let $\xi = [\zeta_3^T, \zeta_2^T]^T$, considering (4), the Eqs. (9) and (10) become

$$\mathcal{D}_1(\xi)\ddot{\xi} + \mathcal{C}_1(\zeta, \dot{\zeta})\dot{\xi} + \mathcal{P}_1 = \mathcal{B}_1\mathcal{U} \tag{11}$$

where

$$U = [-u_1, u_2]^T, \quad \mathcal{D}_1(\zeta) = \begin{bmatrix} \mathcal{L} & \mathcal{J} \\ \mathcal{B} & \mathcal{A} \end{bmatrix}, \quad \mathcal{P}_1 = \begin{bmatrix} \mathcal{K} \\ \mathcal{H} \end{bmatrix}, \quad \mathcal{B}_1 = \begin{bmatrix} D_{31}D_{11}^{-1} & 0 \\ D_{21}D_{11}^{-1} & I \end{bmatrix},$$

$$C_1(\zeta, \dot{\zeta}) = \begin{bmatrix} C_{33} - D_{31}D_{11}^{-1}C_{13} & C_{32} - D_{31}D_{11}^{-1}C_{12} \\ C_{23} - D_{21}D_{11}^{-1}C_{13} & C_{22} - D_{21}D_{11}^{-1}C_{12} \end{bmatrix}.$$

Considering (8) and (11), we have

$$\begin{aligned} \ddot{\zeta}_1 = & -(D_{31}^{-1}\mathcal{J} + D_{11}^{-1}D_{12})\ddot{\zeta}_2 - (D_{31}^{-1}\mathcal{L} + D_{11}^{-1}D_{13})\ddot{\zeta}_3 \\ & - D_{31}^{-1}(\mathcal{F} + \mathcal{K}) - D_{11}^{-1}(C_1 + G_1 + d_{11}) \end{aligned} \tag{12}$$

$$\ddot{\xi} = -D_1^{-1}C_1(\zeta, \dot{\zeta})\dot{\xi} - D_1^{-1}\mathcal{P}_1 + D_1^{-1}\mathcal{B}_1U_1 \tag{13}$$

where D_{31}^{-1} is the inverse of D_{31} .

Property 3 The inertia matrix \mathcal{D}_1 is symmetric and positive definite.

Property 4 The matrix $\dot{\mathcal{D}}_1 - 2C_1$ is skew-symmetric.

Property 5 The eigenvalues of the inertia matrix \mathcal{B}_1 are positive.

Assumption 2.2 There exist known finite positive constants $b_1 > 0$ and $b_2 > 0$ such that $\forall \zeta \in R^n, b_1 \leq \|\mathcal{B}_1\| \leq b_2$.

In the following study, let $\|\cdot\|$ denote the 2-norm, i.e., given $A = [a_{ij}] \in R^{m \times n}$, $\|A\| = \sqrt{\sum_{i=1}^m \sum_{j=1}^n \|a_{ij}\|^2}$.

3 Control Objectives for the Switching Joints

For the switching joints, we give the following assumptions for the actuated and passive modes, respectively,

Assumption 3.1 (Actuated Switching Joints) [6, 7] The desired trajectories $\zeta_{1d}(t), \zeta_{2d}(t), \zeta_{3d}(t)$ and their time derivatives up to the third-order are continuously differentiable and bounded for all $t \geq 0$.

For the switching joints in the actuated mode, we could design the controllers ensuring that the tracking errors for the variables $\zeta_1, \zeta_2, \zeta_3$ from any $(\zeta_j(0), \dot{\zeta}_j(0)) \in \Omega$, where $j = 1, 2, 3, \zeta_j, \dot{\zeta}_j$ converge to a manifold Ω_{ad} specified as $\Omega_{ad} = \{(\zeta_j, \dot{\zeta}_j) \mid |\zeta_j - \zeta_{jd}| \leq \delta_{j1}, |\dot{\zeta}_j - \dot{\zeta}_{jd}| \leq \delta_{j2}\}$, where $\delta_{ji} > 0, i = 1, 2$. Ideally, ϵ_{ji} should be the threshold of measurable noise. At the same time, all the closed-loop signals are to be kept bounded.

Assumption 3.2 (Under-actuated Switching Joints) [6, 7] The desired trajectories $\zeta_{2d}(t)$, $\zeta_{3d}(t)$ and their time derivatives up to the third-order are continuously differentiable and bounded for all $t \geq 0$.

The control objective for the motion of the system with the unactuated switching joint is to design, if possible, controllers that ensure the tracking errors for the variables ζ_2, ζ_3 from any $(\zeta_2(0), \zeta_3(0), \dot{\zeta}_2(0), \dot{\zeta}_3(0)) \in \Omega$, $\zeta_2, \dot{\zeta}_2, \zeta_3, \dot{\zeta}_3$ to converge to a manifold Ω_{ud} specified as Ω where $\Omega_{ud} = \{(\zeta_k, \dot{\zeta}_k) \mid |\zeta_k - \zeta_{kd}| \leq \delta_{k1}, |\dot{\zeta}_k - \dot{\zeta}_{kd}| \leq \delta_{k2}\}$, where $\delta_{ki} > 0, i = 1, 2, k = 2, 3$. Ideally, δ_{ki} should be the threshold of measurable noise. At the same time, all the closed-loop signals are to be kept bounded.

In the following, we can analyze and design the control for each subsystem. For clarity, define the tracking errors and the filtered tracking errors as $e_j = \zeta_j - \zeta_{jd}$, $r_j = \dot{e}_j + \Lambda_j e_j$, where Λ_j is positive definite, $j = 1, 2, 3$. In addition, the following computable signals are defined: $\check{\zeta}_{jr} = \dot{\zeta}_{jd} - \Lambda_j e_j$, $\check{\zeta}_{jr} = \check{\zeta}_{jd} - \Lambda_j \dot{e}_j$.

4 Dynamic Coupling Switching Control

4.1 Actuated Switching Joint

Since $\dot{\zeta}_j = \dot{\zeta}_{jr} + r_j$, $\check{\zeta}_j = \check{\zeta}_{jr} + \dot{r}_j$, where $j = 1, 2, 3$, the Eq. (4) becomes

$$\mathcal{D}\dot{\mathbf{r}} + \mathbf{C}\mathbf{r} = -\mathcal{D}\check{\zeta}_r - \mathcal{C}\check{\zeta}_r - \mathcal{G} - d + \mathbf{U} \tag{14}$$

where $\mathbf{r} = [\mathbf{r}_1^T, \mathbf{r}_2^T, \mathbf{r}_3^T]^T$, $\check{\zeta}_r = [\check{\zeta}_{1r}^T, \check{\zeta}_{2r}^T, \check{\zeta}_{3r}^T]^T$.

The unknown continuous function $-\mathcal{D}\check{\zeta}_r - \mathcal{C}\check{\zeta}_r - \mathcal{G} - d$ in (14) can be approximated by FLSs to arbitrary accuracy as

$$\mathcal{D}\check{\zeta}_r + \mathcal{C}\check{\zeta}_r + \mathcal{G} + d = \mathbf{W}^{*T}S(\mathbf{Z}) + \delta(\mathbf{Z}) \tag{15}$$

where the input vector $\mathbf{Z} = [\check{\zeta}_r, \dot{\zeta}_r, \dot{\zeta}, \zeta, 1]^T \in R^{5n}$. Note that the input vector \mathbf{Z} is composed of real elements (i.e., $\mathbf{Z} \in R^{5n}$). So, if κ fuzzy labels are assigned to each of these elements, the total number of fuzzy rules, L , in the FLS of ζ is κ^5 . For instance, if κ was chosen to be three, the FLS of each of the robots has to fire 243 rules in order to compute and dispatch the control signals to the τ . Moreover, $\delta(\mathbf{Z})$ is the approximation error satisfying $|\delta(\mathbf{Z})| \leq \bar{\delta}$, where $\bar{\delta}$ is an unknown positive constant; \mathbf{W}^* are unknown ideal constant weights satisfying $\|\mathbf{W}^*\| \leq \omega_{\max}$, where ω_{\max} is an unknown positive constant; and $S(\mathbf{Z})$ are the basis functions.

Remark 2 We can obtain an upper bound for the fuzzy logic system, from Eq. (7), it shows $\mathbf{W}^*S(\mathbf{Z}) + \delta(\mathbf{Z}) \leq \|\mathbf{W}^*S(\mathbf{Z})\| + \|\delta(\mathbf{Z})\| \leq \|S(\mathbf{Z})\|\omega_{\max} + \bar{\delta}_{\max} \leq \Theta\Phi(\mathbf{Z})$, where $\Phi(\mathbf{Z}) = \sqrt{\sum_{m=1}^l S^2(\mathbf{Z})} + 1$, and $\Theta = \max\{\bar{\delta}_{\max}, \omega_{\max}\}$.

By using $\hat{\Theta}$ to approximate Θ , let $\Theta = \hat{\Theta} - \tilde{\Theta}$. As Θ is a constant vector, it is easy to obtain that $\dot{\hat{\Theta}} = \dot{\tilde{\Theta}}$. Define control inputs as

$$\mathcal{U} = -K_{p1}\mathbf{r} - \frac{\hat{\Theta}\Phi^2(\mathbf{Z})}{\|\mathbf{r}\|\Phi(\mathbf{Z}) + \beta} \quad (16)$$

$$\dot{\hat{\Theta}} = -\alpha\hat{\Theta} + \gamma\|\mathbf{r}\|\Phi(\mathbf{Z}) \quad (17)$$

where K_{p1} is a diagonal positive constant, $\gamma > 0$, $\beta > 0$, and $\alpha > 0$ are designed parameters and satisfying $\lim_{t \rightarrow \infty} \alpha = 0$, $\int_0^\infty \alpha(s) \, ds = \varrho_\alpha < \infty$, $\lim_{t \rightarrow \infty} \beta = 0$, $\int_0^\infty \beta(s) \, ds = \varrho_\beta < \infty$ with finite constant ϱ_α and ϱ_β .

The stability analysis is omitted here.

4.2 Unactuated Switching Joint

ζ_2 and ζ_3 -subsystems Since $\dot{\xi} = \dot{\xi}_r + r$, $\ddot{\xi} = \ddot{\xi}_r + \dot{r}$, the Eq. (11) becomes

$$\mathcal{D}_1\dot{r} + \mathcal{C}_1r = -\mathcal{D}_1\ddot{\xi}_r - \mathcal{C}_1\dot{\xi}_r - \mathcal{P}_1 + \mathcal{B}_1\mathcal{U} \quad (18)$$

where $r = [r_3^T, r_2^T]^T$, $\ddot{\xi}_r = [\ddot{\xi}_{3r}^T, \ddot{\xi}_{2r}^T]^T$.

The unknown continuous function $\mathcal{D}_1\ddot{\xi}_r + \mathcal{C}_1\dot{\xi}_r + \mathcal{P}_1$ and \mathcal{B}_1^{-1} in (18) can be approximated by FLSs to arbitrary any accuracy as

$$\mathcal{D}_1\ddot{\xi}_r + \mathcal{C}_1\dot{\xi}_r + \mathcal{P}_1 = W_1^*S_1(\mathbf{Z}) + \delta_1(\mathbf{Z}) \quad (19)$$

$$\mathcal{B}_1 - \mathcal{B}_1^0 = W_2^*S_2(\mathbf{Z}) + \delta_2(\mathbf{Z}) \quad (20)$$

Let W_1^0 and W_2^0 be nominal parameter vectors which give the corresponding nominal function $\mathcal{D}_1^0\ddot{\xi}_r + \mathcal{C}_1^0\dot{\xi}_r + \mathcal{P}_1^0$ and \mathcal{B}_1^0 . By Remark 46.2, we have $W_1^*S_1(\mathbf{Z}) + \delta_1(\mathbf{Z}) \leq \|W_1^*S_1(\mathbf{Z})\| + \|\delta_1(\mathbf{Z})\| \leq \|S_1(\mathbf{Z})\|\omega_{1\max} + \delta_{1\max} \leq \Theta_1\Phi_1(\mathbf{Z})$, $W_2^*S_2(\mathbf{Z}) + \delta_2(\mathbf{Z}) \leq \|W_2^*S_2(\mathbf{Z})\| + \|\delta_2(\mathbf{Z})\| \leq \|S_2(\mathbf{Z})\|\omega_{2\max} + \delta_{2\max} \leq \Theta_2\Phi_2(\mathbf{Z})$,

where $\Phi_1(\mathbf{Z}) = \sqrt{\sum_{m=1}^l S_1^2(\mathbf{Z})} + 1$, and $\Theta_1 = \max\{\delta_{1\max}, \omega_{1\max}\}$, $\Phi_2(\mathbf{Z}) = \sqrt{\sum_{m=1}^l S_2^2(\mathbf{Z})} + 1$, and $\Theta_2 = \max\{\delta_{2\max}, \omega_{2\max}\}$.

Define now control new inputs as

$$\mathcal{U} = \mathcal{U}_1 + \mathcal{U}_2 \quad (21)$$

where $\mathcal{U}_1 = (\mathcal{B}_1^0)^{-1} \left(-K_{p2}r - \frac{r\hat{\Theta}_1\Phi_1^2(\mathbf{Z})}{\|\mathbf{r}\|\Phi_1(\mathbf{Z}) + \delta} \right)$, $\dot{\hat{\Theta}}_1 = -\alpha_1\hat{\Theta}_1 + \gamma_1\|\mathbf{r}\|\Phi_1(\mathbf{Z})$, where K_{p2} is diagonal positive constant, \mathcal{U}_2 is designed to compensate for the parameter

errors and the function approximation errors arising from approximating the unknown function as $\mathcal{U}_2 = \frac{1}{b_1} \frac{r\hat{\Theta}_2\Phi_2^2(Z)\|U_1\|^2}{\|r\|\Phi_2(Z)\|\mathcal{U}_1\|+\delta}$, where δ is designed parameter and satisfies $\lim_{t \rightarrow \infty} \delta = 0$, $\int_0^\infty = \varrho_\delta < \infty$ with finite constant ϱ_δ . $\hat{\Theta}_2$ denotes the estimate $\Theta_2(t)$, which is adaptively turned according to $\dot{\hat{\Theta}}_2 = -\alpha_2\hat{\Theta}_2 + \gamma_2\|r\|\|\mathcal{U}_1\| \Phi_2(Z)$, with $\alpha_i > 0$ is design parameters and satisfies $\lim_{t \rightarrow \infty} \alpha_i = 0$, $\int_0^\infty = \varrho_\alpha < \infty$ with finite constant ϱ_α and γ_i , which are design parameters. The stability analysis is omitted here.

ζ_1 -subsystem Finally, for system (5)–(7) under control laws (21), apparently, the ζ_1 -subsystem (5) can be rewritten as $\dot{\varphi} = f(v, \varphi, \mathcal{U})$, where $\varphi = [\zeta_1^T, \dot{\zeta}_1^T]^T$, $v = [r^T, \dot{r}^T]^T$, $\mathcal{U} = [u_1^T, u_2^T]^T$.

Assumption 4.1 From (6) and (7), the reference signal satisfies Assumption 6.4, and the following functions are Lipschitz in γ , i.e., there exists Lipschitz positive constants L_γ and L_f such that $\|C_1 + G_1 + d_1\| \leq L_{1\gamma}\|\gamma\| + L_{1f}$, $\|\mathcal{F} + \mathcal{K}\| \leq L_{2\gamma}\|\gamma\| + L_{2f}$, moreover, from the stability analysis of ζ_2 and ζ_3 subsystems, γ converges to a small neighborhood of $\gamma_d = [\zeta_{3d}, \zeta_{2d}, \dot{\zeta}_{3d}, \dot{\zeta}_{2d}]^T$.

Remark 3 Under the stability of ζ_2 and ζ_3 subsystems, let $\|\gamma - \gamma_d\| \leq \varsigma_1$, it is easy to obtain $\|\gamma\| \leq \|\gamma_d\| + \varsigma_1$, and similarly, let $\mu = [\check{\zeta}_3, \check{\zeta}_2]^T$, and $\mu_d = [\check{\zeta}_{3d}, \check{\zeta}_{2d}]^T$, $\|\mu\| \leq \|\mu_d\| + \varsigma_2$, where ς_1 and ς_2 are small bounded errors.

Lemma 1 [2] The ζ_1 -subsystem (5), if ζ_2 -subsystem and ζ_3 -subsystem are stable, is globally asymptotically stable, too.

Theorem 1 [2] Consider the system (5–7) with Assumptions 6.4, under the action of control laws (21). For compact set Ω , where $(\zeta_2(0), \zeta_3(0), \dot{\zeta}_2(0), \dot{\zeta}_3(0)) \in \Omega$, the tracking errors r converges to the compact sets Ω , and all the signals in the closed-loop system are bounded.

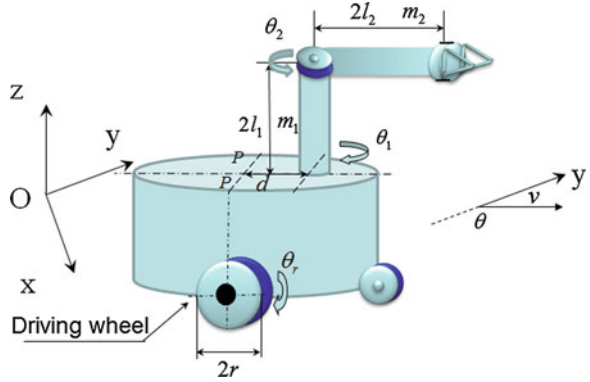
For the system switching stability between the actuated and under-actuated mode, we give the following theorem

Theorem 2 [2] Consider the system (4) with the actuated mode (4) and the under-actuated mode (5–7), if the system is both stable before and after the switching phase using the control laws (16) and (21), and assume that there exists no external impacts during the switching, then the system is also stable during the switching phase.

5 Simulations

Consider a mobile wheeled switching actuated manipulators shown in Fig. 2. The mobile manipulator is subjected to the following constraint: $\dot{x} \cos \theta - \dot{y} \sin \theta = 0$. Using Lagrangian approach, we can obtain the standard form with $q = [\theta_l, \theta_r, \theta_1]^T$,

Fig. 2 The mobile hybrid-actuated manipulator in the simulation



$\dot{\zeta} = [\dot{\zeta}_1, \dot{\zeta}_2, \dot{\zeta}_3]^T = [v, \dot{\theta}, \dot{\theta}_1]^T$, then we could obtain $M(q)\ddot{q} + C(q, \dot{q})\dot{q} + G(q) = B(q)\tau + J^T f$. In the simulation, we assume the parameters $p_0 = 6.0 \text{ kg} \cdot \text{m}^2$, $p_1 = 1.0 \text{ kg} \cdot \text{m}^2$, $p_2 = 0.5 \text{ kg} \cdot \text{m}^2$, $p_3 = 1.0 \text{ kg} \cdot \text{m}^2$, $p_4 = 2.0 \text{ kg} \cdot \text{m}^2$, $q_0 = 4.0 \text{ kg} \cdot \text{m}^2$, $q_1 = 1.0 \text{ kg} \cdot \text{m}^2$, $q_2 = 1.0 \text{ kg} \cdot \text{m}^2$, $q_3 = 1.0 \text{ kg} \cdot \text{m}^2$, $q_4 = 0.5 \text{ kg} \cdot \text{m}^2$, $d = 1.0 \text{ m}$, $r = 0.5 \text{ m}$. The disturbances from environments on the system are introduced as $0.1 \sin(t)$, $0.1 \sin(t)$, and $0.1 \sin(t)$ to the simulation model. The desired trajectories are chosen as $\zeta_1 = 0 \text{ rad}$ and $\theta_d = 0.0 \text{ rad/s}$, $\zeta_2 = 0.5t \text{ m}$ and $v_d = 0.5 \text{ m/s}$, $\theta_{1d} = 0 \text{ rad}$, and the initial value is $[0.0, 0.1, 0.1, -0.2, 0.1, \pi/18]^T$. The design parameters of the controller for full actuated joints are: $A = \text{diag}[1, 1, 1]$, $K = \text{diag}[10, 10, 10]$, 200 training data are sampled for the sliding window during $[0, 5] \text{ s}$ as the phase I. The control $\mathcal{U} = -\frac{1}{2}R_1^{-1} \mathbf{r} + \hat{\mathbf{U}}_r$. From $t = [5, 10] \text{ s}$ as the phase II, the switching joint is switched to passive mode, the parameters of the fuzzy controller for under-actuated joints are:

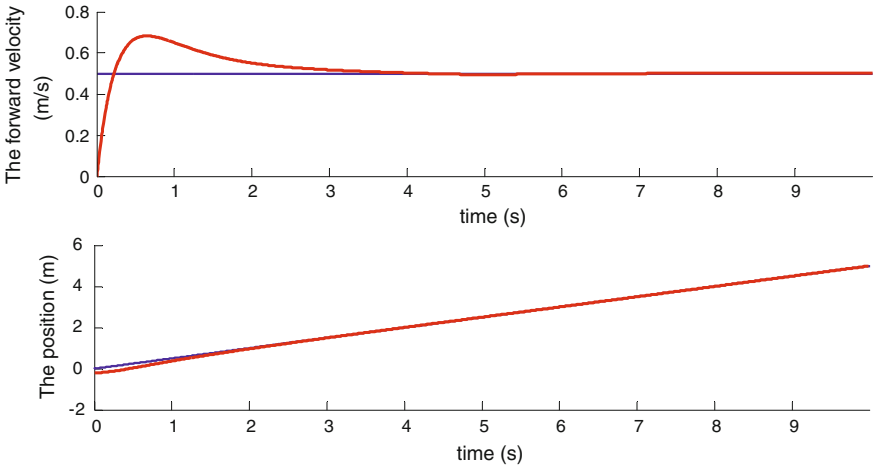


Fig. 3 Tracking the forward velocity v and the corresponding ζ_1

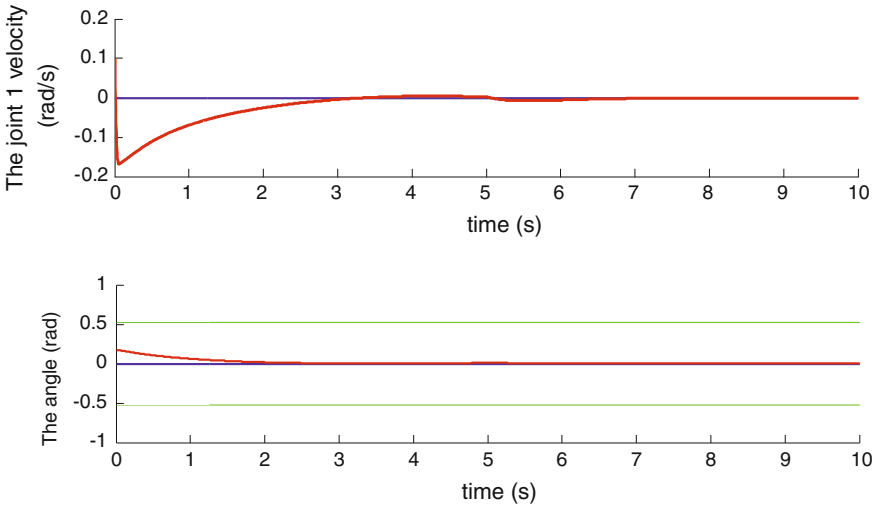


Fig. 4 Tracking the desired position of θ_1

$A = \text{diag}[1, 1]$, $K = \text{diag}[10, 10]$. The control $U = -\frac{1}{2}R_2^{-1}r + \hat{U}_f$. From $t = [10, 15]$ s as the phase III, the switching joint is switched to the original mode. The used controller is the same as the phase I. The trajectory profiles for ζ_1 and ζ_1 are shown in Fig. 3. The positions tracking and the corresponding velocity profiles for the ζ_2 and ζ_3 are shown in Figs. 4 and 5 during the whole switching, the switching points happen on $t = 5$ and 10 s, we can see the positions and velocities are continuous and converge to the desired values, which is stable and bounded during the switching.

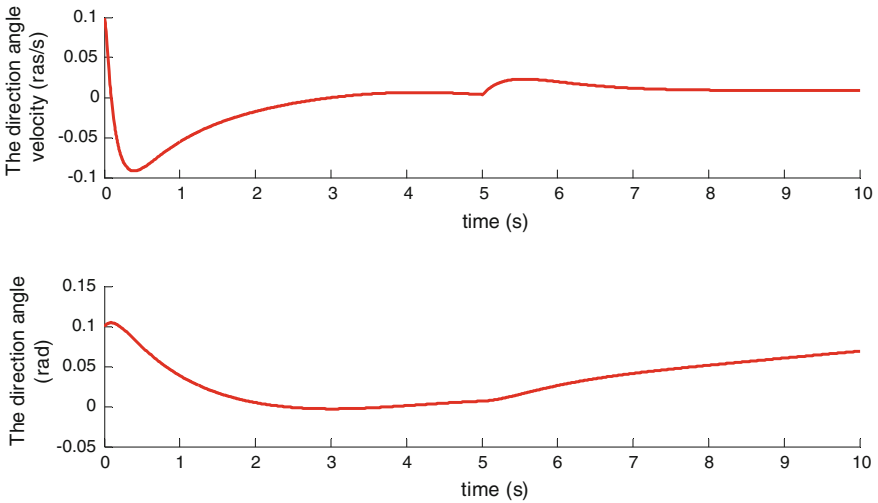


Fig. 5 The profiles of $\zeta_2 = \theta$ and ω

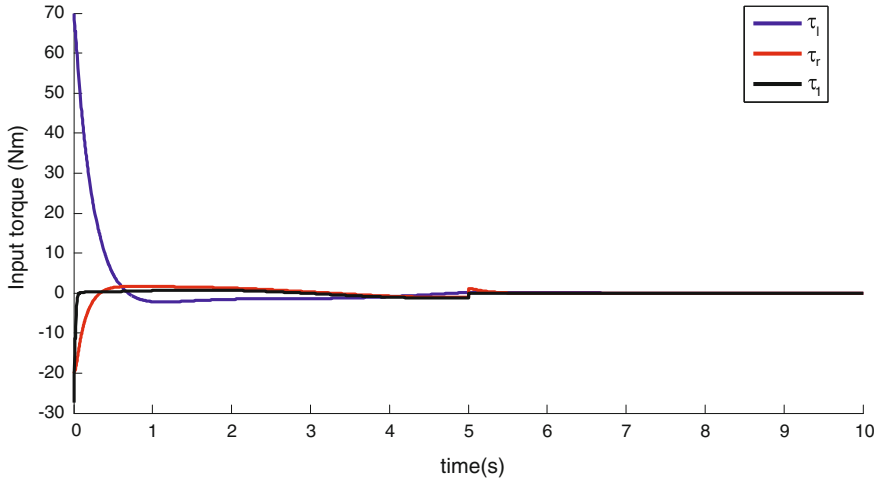


Fig. 6 Input torques

The input torques of the pre-switch and post-switch are shown in Fig. 6.

6 Conclusion

In this paper, adaptive fuzzy switching control designs are carried out for dynamic balance and tracking of desired trajectories of mobile manipulators with switching actuated joints in the presence of unmodelled dynamics, or parametric/functional uncertainties. Simulation results demonstrate that the system is able to track reference signals satisfactorily, with all closed-loop signals uniformly bounded.

Acknowledgments This work is supported in part by the Natural Science Foundation of China under Grant 61174045 and Grant 61111130208, in part by the International Science and Technology Cooperation Program of China under Grant 2011DFA10950, and in part by the Fundamental Research Funds for the Central Universities under Grant 2011ZZ0104 the Program for New Century Excellent Talents in University No. NCET-12-0195.

References

1. Li Z, Ming A, Xi N, Gu J, Shimojo M (2005) Development of hybrid joints for the complaint arm of human-symbiotic mobile manipulator. *Int J Robot Autom* 20(4):260–270
2. Li Z, Kang Y (2010) Dynamic coupling switching control incorporating support vector machines for wheeled mobile manipulators with hybrid joints. *Automatica* 46(5):785–958
3. Li Z, Cao X, Ding N (2011) Adaptive fuzzy control for synchronization of nonlinear teleoperators with stochastic time-varying communication delays. *IEEE Trans Fuzzy Syst* 19(4):745–757

4. Li Z, Xu C (2009) Adaptive fuzzy logic control of dynamic balance and motion for wheeled inverted pendulums. *Fuzzy Sets Syst* 160(12):1787–1803
5. Li Z, Gu J, Ming A, Xu C (2006) Intelligent compliant force/motion control of nonholonomic mobile manipulator working on the non-rigid surface. *Neural Comput Appl* 15(3–4):204–216
6. Chang YC, Chen BS (2000) Robust tracking designs for both holonomic and nonholonomic constrained mechanical systems: adaptive fuzzy approach. *IEEE Trans Fuzzy Syst* 8:46–66
7. Li Z, Yang C, Fan L (2012) *Advanced control of wheeled inverted pendulum systems*. Springer, London

Bundle Branch Blocks Classification Via ECG Using MLP Neural Networks

Javier F. Fornari and José I. Peláez Sánchez

Abstract This paper proposes a two-stage system based on neural network models to classify bundle branch blocks via electrocardiogram (ECG) analysis. Two artificial neural network (ANN) models have been developed in order to discriminate bundle branch blocks and hemiblocks from normal ECG and other heart diseases. This method includes pre-processing and classification modules. ECG segmentation and wavelet transform were used as pre-processing stage to improve classical multilayer perceptron (MLP) network. A new set of about 800 ECG were collected from different clinics in order to create a new ECG database to train ANN models. For bundle branch blocks classifier in the test phases, the best specificity of all models was found to be 94.56 % and the best sensitivity was found to be 92.45 %. In the case of hemiblocks classifier, the best results were a sensitivity of 93.26 % and a specificity of 92.55 %.

Keywords Classification · ECG · Bundle branch blocks · Hemiblocks · MLP · DWT

1 Introduction

Bundle branch blocks are diseases related to defects in the heart's electrical conduction system. This system begins in the sinoatrial node, which acts as the heart's natural pacemaker, situated on the upper right atrium. As shown in Fig. 1, the heart's electrical impulse travels through the atria by several roads converging

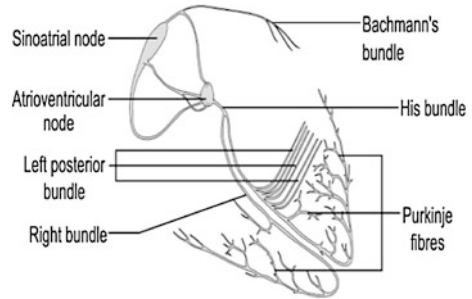
J. F. Fornari (✉)

Argentinian Technological University, Rafaela 2300 Santa Fe, Argentina
e-mail: javier.fornari@fra.utn.edu.ar

J. I. Peláez Sánchez

Department of Languages and Computer Sciences, Malaga University, Malaga, Spain
e-mail: jipelaez@uma.es

Fig. 1 Heart's electrical impulse



in the atrioventricular (AV) node. In the AV node, the electrical impulse is delayed in order to allow full contraction of atria, and then, it travels down the His bundle and splits into the left and right bundle branches.

The left bundle branch subdivides into the left anterior fascicle and the left posterior fascicle (later divided into two again, with one fascicle being the septal fascicle), while the right bundle branch contains only one fascicle. All the fascicles divide into Purkinje fibers, which in turn interdigitize with individual cardiac myocytes, allowing for physiologic depolarization of the ventricles.

Several reasons may cause damage to the bundle branches or fascicles that impair their ability to transmit electrical impulses appropriately, such as underlying heart disease, myocardial infarction, and so on. When this happens, the electrical impulse may be delayed (first degree), intermittently stopped (second degree), or completely stopped (third degree), altering the pathways for ventricular depolarization. Whether the electrical impulses change its course, it can cause delays and changes in the directional propagation of the impulses. These disturbances cause a loss of ventricular synchrony and a drop in cardiac output. A pacemaker may be required to restore an optimal electrical supply to the heart.

Since electrocardiography expresses heart's electrical activity, heart diseases could be diagnosed by morphological study of recorded data. Cardiologist commonly used this technique since it consists of effective, non-invasive, and low-cost tool to the diagnosis of cardiovascular diseases. For this purpose, ECG record is made to examine and observe a patient.

The most representative sign of bundle branch blocks and hemiblocks lies in the QRS complex. Left bundle branch blocks (LBBB) widen the entire QRS and generally shift the heart's electrical axis to the left. Meanwhile, right bundle branch blocks (RBBB) widen only the last part of the QRS complex and may shift the heart's electrical axis slightly to the right.

Furthermore, first-degree bundle branch blocks widen QRS complex in ECG due to the delay introduced in any point of the fascicle. As shown in Fig. 2, RBBB affects mainly the last part of QRS complex in all eight leads (dotted line), compared with healthy patients (solid line). The other characteristic features of the signal remain practically unchanged [1].

In this study, the standard 12-lead ECG has been used in order to record the clinical information of each patient. The term "lead" refers to the tracing of the

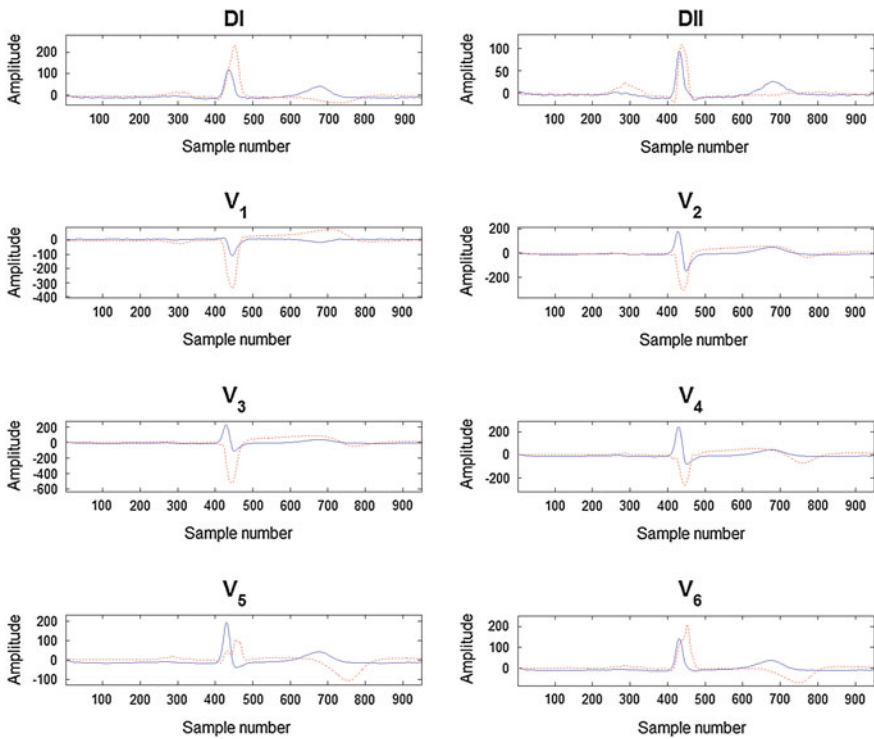


Fig. 2 Healthy patient versus patient with *left* bundle branch block

voltage difference between two of the electrodes, namely I, II, III, aVL, aVR, aVF, V1, V2, V3, V4, V5, and V6. However, the first six leads are a linear combination, so only two of them are enough to gather all necessary information for analysis [2].

There are many approaches for computer processing of ECG for diagnosing certain heart diseases. The two mainly used strategies are methods based on morphological analysis [3–5] and methods based on statistical models [6–8]. A third group, methods based on artificial neural networks (ANNs) [9–11], is being developed lately focusing on ECG signal classification.

Several techniques have been applied to ECG for feature extraction, such as discrete cosine transform (DCT) [12], discrete Fourier transform (DFT) [13], continuous wavelet transform (CWT) [14], discrete wavelet transform (DWT) [15], principal component analysis (PCA) [16], and multidimensional component analysis (MCA), among others.

ANN has been used in a wide variety of applications, such as classification tasks or pattern recognition. Multilayer perceptron (MLP) is a traditional ANN model, in which each neuron computes the weighted sum of its inputs and applies to sum of a non-linear function called activation function. The performance of MLP depends mainly on the learning algorithm, the number of hidden layers, the number of hidden neurons, and the activation function for each neuron. The

commonly investigated activation function in the literature is sigmoid function, which is fixed and cannot be adjusted to adapt to different problems, but it is critical in the performance of MLP. This function represents the neuron response: a relationship between a single input, the weighted sum, and a single output. MLP and radial basis function (RBF) have shown very good learning and predicting capabilities in the classification question. [17–19]

In this study, we propose a two-stage system for bundle branch blocks detection. The first stage is responsible for signal enhancement and feature extraction using the signal averaging and wavelet transform [20]. For the classification stage, we opted for the MLP.

This paper is divided into six sections, following this initial introductory section (Sect. 1), the basis of the employed signal processing and neural network will be commented (Sect. 2), and the most important details of the new ECG database used will be discussed (Sect. 3). Next, the pre-processing stage (Sect. 4) and classification stage (Sect. 5) will be described. And finally, the results (Sect. 6) and conclusions (Sect. 7) of the study will be explained.

2 Background

2.1 Continuous Wavelet Transform

Since 1982, when Jean Morlet proposed the idea of the wavelet transform, many people have applied the wavelet to different fields, such as noise suppression, image compression, or molecular dynamics. Wavelets are a family of functions generated from translations and dilatations of a fixed function called the “mother wavelet”. The wavelet transform can be thought of as an extension of classic Fourier transform, but it works on a multiscale basis (time and frequency). The wavelet transform can be classified as continuous or discrete. Many researchers (Daubechies, Haar, Meyer, Mallat, etc.) enhanced and developed this signal-processing tool to make it more efficient [21].

The wavelet analysis has been introduced as a windowing technique with variable-sized regions. Wavelet transforms may be considered forms of time–frequency representation for continuous-time signals and introduce the notion of scale as an alternative to frequency, mapping a signal into a time–scale plane. This is equivalent to the time–frequency plane used in the short-time Fourier transform (STFT). Each scale in the time–scale plane corresponds to a certain range of frequencies in the time–frequency plane. A wavelet is a waveform of limited duration. Wavelets are localized waves that extend for a finite time duration compared to sine waves that extend from minus to plus infinity. The wavelet analysis is the decomposition of a signal into shifted and scaled versions of the original wavelet, whereas the Fourier analysis is the decomposition of a signal into sine and cosine waves of different frequencies. The wavelets forming a CWT are subject to the uncertainty principle of Fourier analysis respective sampling theory,

so one cannot assign simultaneously an exact time and frequency response scale to an event.

Mathematically, the CWT of a function is defined as the integral transform with a family of wavelet functions: In other words, the CWT is defined as the sum of the signal multiplied by scaled and shifted versions of the wavelet function. A given signal of finite energy is projected on a continuous family of frequency bands of the form $[f, 2f]$ for all positive frequencies. These frequency bands are scaled versions of a subspace at scale 1. This subspace is in most situations generated by the shifts of the mother wavelet $\Psi(x)$. The projection of a function $f(t)$ onto the subspace of scale a has the form:

$$W_f(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t)\Psi\left(\frac{t-b}{a}\right)dt \tag{1}$$

For the CWT, the pair (a, b) varies over the full half-plane, while for the DWT this pair varies over a discrete subset of it.

2.2 Discrete Wavelet Transform

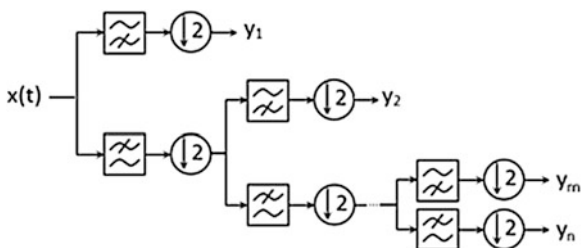
The DWT captures both frequency and time information as the CWT does, but using wavelets discretely sampled. The fast wavelet transform (FWT) is an alternative to the conventional fast Fourier transform (FFT) because it can be performed in $O(n)$ operations and it captures as the notion of frequency content of the input (different scales— a) and as the notion of temporal content (different positions— b).

As shown in Fig. 3, the DWT can be implemented by a series of filters. First, the samples $(s[n])$ are passed through a low-pass filter with impulse response $l[n]$ resulting in a convolution of the two, and simultaneously, the samples are passed through a high-pass filter with impulse response $h[n]$ (Eq. 2).

$$Y_{low}[n] = (s * l)[n] = \sum_{k=-\infty}^{\infty} s[k]l[n - k] \tag{2}$$

$$Y_{high}[n] = (s * h)[n] = \sum_{k=-\infty}^{\infty} s[k]h[n - k]$$

Fig. 3 Filter bank representation



According to Nyquist's theorem, half of the frequencies of signal have now been removed, so half the samples can be discarded downsampling by 2 the outputs of the filters.

Calculating wavelet coefficients at every possible scale would waste computation time, and it generates a too large volume of data. However, only a subset of scales and positions are needed; scales and positions are based on powers of two, so-called dyadic scales and positions. Then, the wavelet coefficients c_{jk} are given by Eq. 3.

$$c_{jk} = W_f(2^{-j}, k2^{-j}) \quad (3)$$

where $a = 2^{-j}$, is the dyadic dilation and $b = k2^{-j}$, is the dyadic position.

Such an analysis from the DWT is obtained. DWT works like a band-pass filter and DWT for a signal several levels can be calculated. Each level decomposes the input signal into approximations (low-frequency part of initial signal) and details (high-frequency part of initial signal). The next level of DWT is done upon approximations. y_1 corresponds to the first level of DWT (approximations—low-pass filter), y_2 corresponds to the second level of DWT (approximations from details of first level), and so on. The penultimate level (y_m) corresponds to the approximations of the last filter pair, and the last level (y_n) corresponds to the details of the last filter pair [22].

3 Database

A new ECG database was created by Gem-Med, S.L. The Gem-Med database contains eight lead ECG signals of about 800 patients. It is possible to recuperate all the 12 leads from the datasets recorded (I and II leads and the six precordial leads V1 to V6) [2]. These ECG signals are sampled at a frequency of 1,000 Hz and filtrated with a band-pass filter of 0.5 and 45 Hz to correct the baseline and to suppress interferences (motion artifact, power line interference, etc.) [23]. Among the ECG that composed the new database, there are several diagnoses such as healthy patient, different types of bundle branch blocks (RBBB, LBBB, and hemiblocks), different types of ischemias (inferior, lateral, superior, anterior, septal, chronic, acute, etc.), and so on.

Cardiologists have diagnosed each pathology with four possible results, namely Sure, Probable, Discarding, or Negative. Those ECGs showing unequivocal signs of the disease under study are marked as Sure. If a marker was not clear, or does not appear, the ECG is marked as Probable. When there is only a sign but cannot guarantee the diagnosis, the ECG is marked as Discarding. Finally, ECGs that do not show any sign associated with the pathology under study are marked as Negative.

The original data are saved in SCP-ECG format, which stands for Standard Communications Protocol for Computer Assisted Electrocardiography. The ECG signals for training dataset contain eight lead of patients diagnosed with bundle

branch blocks (Sure and Probable degrees) and those who were not diagnosed with this pathology. This dataset was divided into three groups of training, validation, and testing. Each ECG database record is composed of several beats and all the eight leads. After the pre-processing block, the mean beat has been calculated and the number of final variables has been reduced to about 20 samples per lead. Each subnet works with a different pre-processing, so each one manages a different number of samples as input. The input data store all the required leads concatenated in a one-dimensional vector.

4 Signal Pre-processing

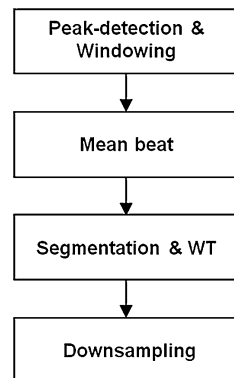
The ECG signal pre-processing (Fig. 4) is performed to remove noise distortion (mean beat), to extract main features (wavelet transform) [24], and to data reduction (downsampling).

Firstly, as the QRS complex is the most prominent wave in ECG, R wave is recognized in each beat with a peak-detection algorithm. A window of 950 samples centered in R wave is defined (Fig. 5). Except in cases of severe bradycardia or tachycardia, complete heart cycle falls within that window and all the significant points are approximately at the same sample number. Locating R wave at the center of the exploration window is achieved when variations in heart rate increased or decreased the number of isoelectric line samples at the edges of the exploration window.

When all the beats are recognized, those that differ more are discarded (artifacts, extra systoles, etc.) and the rest are averaged in order to remove noise distortion. As in this work the diseases targeted are different kinds of bundle branch blocks of first and third degree, beat average does not suppress any relevant information as in other heart diseases like arrhythmias or some bundle branch blocks.

Prognostic factors for bundle branch blocks are mainly located in the QRS complex [1], before the feature extraction stage proceeds to segment the ECG signal. As mentioned previously, except in cases of severe bradycardia or

Fig. 4 Pre-process module



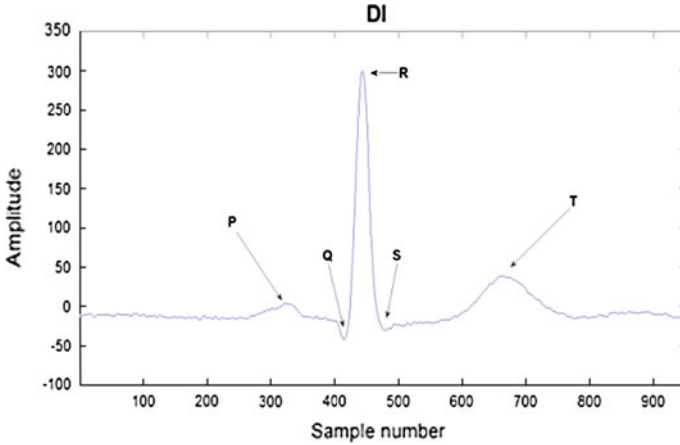


Fig. 5 ECG significant points

tachycardia, by placing the R wave at a specific position, the QRS complex and T wave maintain its lengths and can be extracted by windowing technique.

In this work, the wavelet transform was selected in order to reduce the remaining number of samples. It is also essential to notice that determination of DWT level and the mother wavelet are very important in ECG feature extraction.

In this study, the best results of ECG signals classification were obtained for DWT–MLP structure by examining different already most used mother wavelets to select the best for DWT techniques. Also, optimal decomposition-level parameters in DWT were specified empirically. Further, we compared six different mother wavelets, namely Haar, Daubechies 2, Daubechies 4, Coiflet 1, Symlet 2, and Symlet 4, and a non-DWT process.

The Haar wavelet (Fig. 6) is the simplest possible wavelet, proposed in 1909 by Alfred Haar. Because this wavelet is not continuous, it offers an advantage for detecting sudden transitions, such as the QRS complex in ECG. The Haar wavelet’s mother wavelet function can be described as Eq. 4.

$$\Psi(t) = \begin{cases} 1, & 0 \leq t < \frac{1}{2} \\ -1, & \frac{1}{2} \leq t < 1 \\ 0, & 0 > t \geq 1 \end{cases} \quad (4)$$

The Daubechies wavelets (Fig. 7), named after her inventor Ingrid Daubechies, are a family of orthogonal wavelets characterized by a maximal number of vanishing moments. This type of wavelet is easy to put into practice using the FWT, but it is not possible to write down in closed form.

Symlets (Fig. 8) are also known as the Daubechies least asymmetric wavelets, and their construction is very similar to the Daubechies wavelets. Daubechies

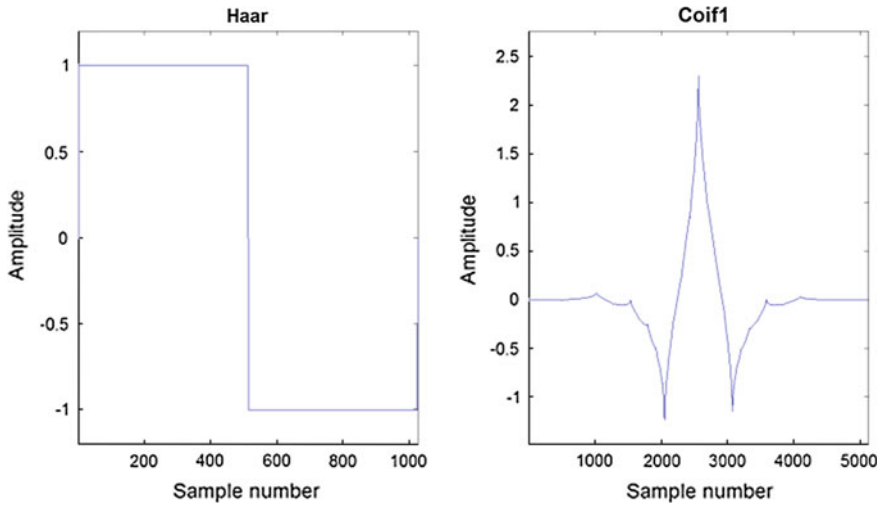


Fig. 6 Haar and Coiflet 1 mother wavelets

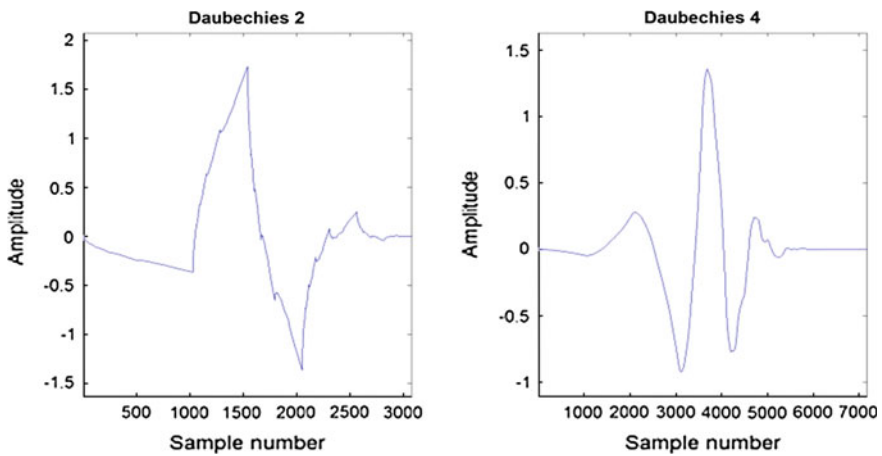


Fig. 7 Daubechies 2 and 4 mother wavelets

proposed modifications of her wavelets that increase their symmetry while retaining great simplicity.

Coiflets (Fig. 6) are a family of orthogonal wavelets designed by Ingrid Daubechies to have better symmetry than the Daubechies wavelets.

Finally, the pre-processing section concludes reducing the number of variables remaining by the first level of wavelet transform. In the case of the subnet without integral transformation, simply the signal is downsampled.

The DWT of an ECG signal is calculated by passing it through a series of filters related by pairs (quadrature mirror filter), being decomposed simultaneously with a

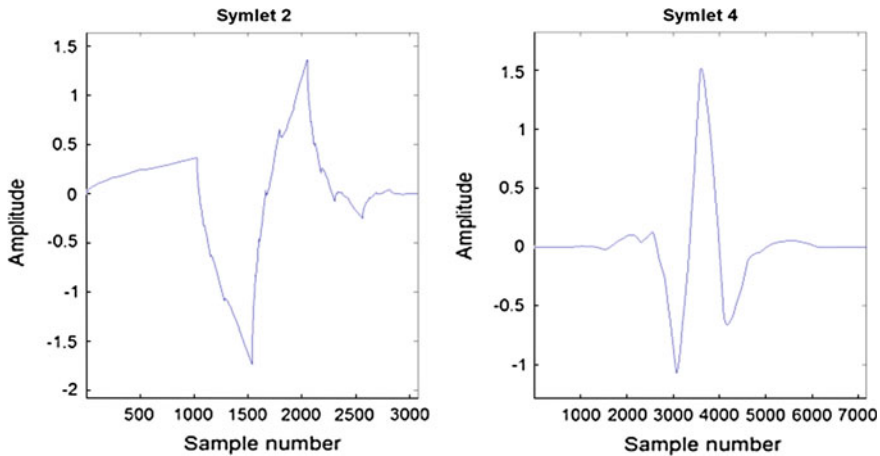


Fig. 8 Symlet 2 and 4 mother wavelets

low-pass filter (approximation coefficients) and with a high-pass filter (details coefficients), and later downsampled by 2. In this study, the wavelet filter outputs selected as inputs to the next stage (MLP) are the approximation coefficients of the very first level.

As the Nyquist's sampling theorem states: "If a function $x(t)$ contains no frequencies higher than B hertz, it is completely determined by giving its ordinates at a series of points spaced $1/(2B)$ seconds apart". Thus, as the original ECG is sampled at a frequency of 1,000 Hz and later is filtered by a band pass between 0.5 and 45 Hz, the information is contained up to 45 Hz and the signal is sampled with an oversampling factor of about 22. This means that the signal can be downsampled by a factor up to 22 without losing significant information. In this study, the downsampling factor selected reduces the number of samples from about 100 to 20 samples (about 5) in the Simple-MLP net.

5 MLP Implementation

In this study, we have designed two system focused on a different family of bundle branch blocks: The first family is composed by right and LBBB and the second family is composed by hemiblocks. For each system, seven different MLPs have been tested, one for each selected ECG signals pre-processing. These ANNs are Haar-MLP, D2-MLP, D4-MLP, COIF1-MLP, SYM2-MLP, SYM4-MLP, and Simple-MLP (with no DWT).

The basic MLP used in this study is formed by three layers: the input layer (with as many neurons as inputs having the net, generally about 20 neurons), one hidden layer (it has been estimated empirically the optimal number of neurons for each ANN—Table 1), and the output layer (which consists of a single neuron). In

Table 1 Results of number of neurons in hidden layer experiment

N°	Sens max (%)	Sens mean (%)	Sens min (%)	Spec max (%)	Spec mean (%)	Spec min (%)
14	94.23	89.56	82.69	97.18	94.71	90.40
15	91.35	87.64	83.65	96.05	94.39	90.96
16	93.27	89.70	86.54	96.05	93.99	90.40
17	95.19	89.29	81.73	98.31	94.71	91.53
18	94.23	89.84	85.58	98.31	94.27	89.27

this study, 20 ANNs were trained for each number of neurons in hidden layer in order to select the best one (20 nets for 50 different number of neurons in hidden layer).

In Table 1 are shown the results obtained in the number of neurons in hidden layer experiment (only results of 14 neurons to 18 neurons are shown, but results of 1 neuron to 50 neurons have been proved). The optimal number of neurons is not associated with the best value of sensitivity or specificity separately, but the compromise between the two factors. So, 18 neurons in the hidden layer were selected for our system.

Therefore, the training algorithm has to calculate the value of the weights that allow the network to correctly classify the ECG. The selected training algorithm was the scaled conjugate gradient back-propagation [25]. This method is a supervised learning method and is a generalization of the delta rule. Since each ECG was previously diagnosed by a team of cardiologists, using a supervised method is a good choice. The scaled conjugate gradient back-propagation algorithm is based on conjugate directions, but this algorithm does not perform a line search at each iteration. Training stops when performance gradient falls below $1e-5$ or validation performance has increased more than six times since the last time it decreased. The initial weights values were set randomly at the beginning of each training. The convergence of the algorithm is achieved by minimizing the root mean square error (known also as the cost function) between the estimated value and the real value. The minimization of the cost function is achieved through the optimization of the weights.

6 Results and Discussion

6.1 Training Results

Training data of ECG bundle branch blocks used in this study were taken from Gem-Med database. Training patterns had been originally sampled at 1,000 Hz before any were filtered, so they were arranged as about 950 samples in the intervals of R–R for all diagnostic, which are called as a complete window. Firstly, the segment selection (QRS complex and T wave for this study) produces an initial reduction in the number of samples of about 400.

Training patterns were formed in mixed order from the ECG pre-processed. The size of the training patterns is 8 leads of about 20 samples for each system, so 160 samples are present. The combination of these training patterns was called as training set.

The optimum number of hidden nodes and learning rate were experimentally determined for each ANN structure. After the proposed structures were trained by the training set, they were tested for healthy ECG. For the stopping criterion of all the networks, maximum number of iterations was set to 10,000 and the desired error value (MSE) was set to 0.001. MLP models have been trained with the training data including a predetermined number of leads depending on the pathology targeted, and the architectures that can produce best results have been determined with trial and error. The test was implemented using ECG records taken from about 800 patients. The ECG records were collected by Gem-Med S.L., Barcelona, Spain.

6.2 Test Results

The results of classification for the right and LBBB detector are shown in Table 2, and those for the hemiblocks detector are shown in Table 3. Each row of the table represents the instances in a specific ANN.

The first group of columns represents the statistical measures of the performance of the ANN with the training set: sensitivity (Sens), specificity (Spec), and the Matthews correlation coefficient (MC). The second group of columns represents the same quality parameters of the ANN with the test set. The statistical measures of the performance of the binary classification test used in this study are the following [26, 27].

Sens column represents sensitivity of the system, related to the system's ability to identify pathological patients. A high value means there are few pathological cases that are not detected by the system. This can also be written as Eq. 5.

$$\text{Sens} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (5)$$

Table 2 Right and left bundle branch blocks MLP results

Subnet	Training set			Test set		
	Sens (%)	Spec (%)	MC (%)	Sens (%)	Spec (%)	MC (%)
No WT	92.11	96.90	89.51	91.82	91.39	76.22
Haar	91.45	96.90	88.98	89.31	92.75	76.89
D2	92.76	96.90	90.04	92.45	94.56	82.62
D4	92.76	96.51	89.52	92.45	92.15	78.01
Coif1	94.08	93.41	86.63	93.08	91.99	78.19
Sym2	88.16	98.45	88.52	85.53	93.50	75.64
Sym4	93.42	94.96	88.04	93.08	91.24	76.86

Table 3 Hemiblocks MLP results

Subnet	Training set			Test set		
	Sens (%)	Spec (%)	MC (%)	Sens (%)	Spec (%)	MC (%)
No WT	97.80	96.13	93.14	96.81	89.82	68.83
Haar	94.51	96.77	91.28	94.68	89.41	66.66
D2	95.60	99.35	95.64	94.68	92.16	72.31
D4	94.51	97.42	92.14	92.55	93.26	73.50
Coif1	95.60	98.71	94.76	93.62	91.06	69.24
Sym2	95.60	95.48	90.50	94.68	90.51	68.81
Sym4	97.80	94.84	91.52	95.74	89.41	67.36

Spec column represents specificity of the system, related to the system’s ability to identify healthy ECG. A high value means there are few healthy cases that are marked as pathological by the system. This can also be written as Eq. 6.

$$Spec = \frac{TN}{TN + FN} \tag{6}$$

MC column represents the Matthews correlation coefficient, which is used as a measure of the quality of binary classifications (two-class—pathological and non-pathological ECG). A high value means that both sensitivity and specificity are high and the system is very reliable. This can also be written as Eq. 7.

$$MC = \frac{(TP * TN - FP * FN)}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}} \tag{7}$$

7 Conclusion

In this work, we have presented two-stage MLP model for ECG signal diagnosis. We focused on two main bundle branch blocks diagnosis groups, namely right and LBBB, and hemiblocks. The automatic detector of bundle branch blocks consists of two stages, namely feature extraction and classifier. The first stage has been implemented with a segmentation module, which is responsible for selecting the section of interest of the ECG, and a pre-processing module, which performs the ECG feature extraction itself. For this work, several wavelet transforms have been compared. The second stage is implemented with MLP, since there are several different WTs, and several MLPs had to be trained, one for each.

The best single net of the first system is D2-MLP, which has been able to differentiate healthy patients by diagnosis of right and LBBB with 92.45 % of sensitivity and 94.56 % of specificity. The best single net of the second system is D4-MLP, which has been able to differentiate healthy patients by diagnosis of hemiblocks with 92.55 % of sensitivity and 93.26 % of specificity.

Each trained subnet has a slight preference for certain electrocardiographic changes in the pathologies under study. So, superimposing the responses of several of them can achieve a significant reduction in undiagnosed cases correctly.

These results are very promising and encourage us to extend this study to other heart diseases and to investigate other pre-processing and post-processing stages.

Acknowledgments This work is supported by the Ministry of Industry, Tourism, and Trade of Spain (Project TSI-020302-2010-136).

References

1. Bayès de Luna, Antoni. Bases de la (2006) electrocardiografía. Semiología electrocardiográfica II: Patrones diagnósticos de crecimiento, bloques y preexcitación. Barcelona : Prous Science, B-15239-06
2. Klabunde RE (2011) Cardiovascular physiology concepts. Lippincott Williams & Wilkins, Philadelphia. ISBN 9781451113846
3. Taouli SA, Bereksi-Reguig F (2010) Noise and baseline wandering suppression of ECG signals by morphological filter. *J Med Eng Technol* 34:87–96
4. Christov I, y otros (2006) Comparative study of morphological and time-frequency ECG descriptors for heartbeat classification. *Med Eng Phys* 28:876–887
5. Sun Y, Chan KL, Krishnan SM (2002) ECG signal conditioning by morphological filtering. *Comput Biol Med* 32:465–479
6. Wiggins M, y otros (2008) Evolving a Bayesian classifier for ECG-based age classification in medical applications. *Appl Soft Comput* 8:599–608
7. Sharma LN, Dandapat S, Mahanta A (2010) ECG signal denoising using higher order statistics in wavelet subbands. *Biomed Sig Process Control* 5:214–222
8. Morise AP, y otros (1992) Comparison of logistic regression and Bayesian-based algorithms to estimate posttest probability in patients with suspected coronary artery disease undergoing exercise ECG. *J Electrocardiol* 25:89–99
9. Gholam H, Luo D, Reynolds KJ (2006) The comparison of different feed forward neural network architectures for ECG signals diagnosis. *Med Eng Phys* 28:372–378
10. Sekkal M, Chick MA, Settouti N (2011) Evolving neural networks using a genetic algorithm for heartbeat classification. *J Med Eng Technol* 35:215–223
11. Moavenian M, Khorrami H (2010) A qualitative comparison of ANN and SVM in ECG arrhythmias classification. *Expert Syst Appl* 37:3088–3093
12. Ahmed SM, y otros (2009) ECG signal compression using combined modified discrete cosine and discrete wavelet transforms. *J Med Eng Technol* 33:1–8
13. Mitra S, Mitra M, Chaudhuri BB (2004) Generation of digital time database from paper ECG records and Fourier transform-based analysis for disease identification. *Comput Biol Med* 34:551–560
14. Khorrami H, Moavenian M (2010) A comparative study of DWT, CWT, and DCT transformations in ECG arrhythmias classification. *Expert Syst Appl* 37:5151–5157
15. Ranjith P, Baby PC, Joseph P (2003) ECG analysis using wavelet transform: application to myocardial ischemia detection. *ITBM RBM* 24:44–47
16. Ceylan R, Ozbay Y (2007) Comparison of FCM, PCA and WT techniques for classification ECG arrhythmias using artificial neural network. *Expert Syst Appl* 2007:286–295
17. Gaetano A, y otros (2009) A patient adaptable ECG beat classifier based on neural networks. *Appl Math Comput* 213:243–249
18. Ozbay Y, Tezel G (2010) A new method for classification of ECG arrhythmias using neural network with adaptive activation function. *Digit Sig Process* 20:1040–1049

19. Korurek M, Dogan B (2010) ECG beat classification using swarm optimization and radial basis function neural network. *Expert Syst Appl* 37:7563–7569
20. Lin CC (2008) Enhancement of accuracy and reproducibility of parametric modeling for estimating abnormal intra-QRS potentials in signal-averaged electrocardiograms. *Med Eng Phys* 30:834–842
21. Daubechies I (1992) Ten lectures on wavelets. Society for Industrial and Applied Mathematics, Philadelphia. ISBN 0-89871-274-2
22. Vetterli M, Cormac H (1992) Wavelets and filter banks: theory and design. *IEEE Trans Sig Process* 4:2207–2232
23. Mak JNF, Hu Y, Luk KDK (2010) An automated ECG-artifact removal method for trunk muscle surface EMG recordings. *Med Eng Phys* 32:840–848
24. Li C, Zheng C, Tai C (1995) Detection of ECG characteristic points using wavelet transform. *IEEE Trans Biomed Eng* 42:21–28
25. Møller MF (1993) A scaled conjugate gradient algorithm for fast supervised learning. *Neural Netw* 6:525–533
26. Baldi P, y otros (2000) Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinf Rev* 16:412–424
27. Ennett CM, Frize M, Charette E (2004) Improvement and automation of artificial neural networks to estimate medical outcomes. *Med Eng Phy* 26:321–328

Fuzzy Associative Classifier for Probabilistic Numerical Data

Bin Pei, Tingting Zhao, Suyun Zhao and Hong Chen

Abstract Recently, a number of advanced data collection and processing methodologies have led to the proliferation of uncertain data. When discovering from such uncertain data, we should handle these uncertainties with caution, because classical mining algorithms may not be appropriate for uncertain tasks. This paper proposes a generic framework of fuzzy associative classifier for probabilistic numerical data, which is prevalent in the real-world applications, such as sensor networks and GPS-based location. In this paper, we first introduce an Apriori-based algorithm for mining fuzzy association rules from a probabilistic numerical dataset based on novel support and confidence measures suitable for such dataset. Then, we give fuzzy rules redundancy pruning strategy and database coverage method to build a compact fuzzy associative classifier in removing redundant rules and thus improving the accuracy of the classifier. We also redefine multiple fuzzy rules classification method for classifying new instances. Extensive experimental results show the effectiveness and efficiency of our algorithm.

B. Pei (✉) · T. Zhao · S. Zhao · H. Chen

Key Laboratory of Data Engineering and Knowledge Engineering, MOE, Beijing,
People's Republic of China
e-mail: pei_ice@ruc.edu.cn

T. Zhao

e-mail: zhaotingting@ruc.edu.cn

S. Zhao

e-mail: zhaosuyun@ruc.edu.cn

H. Chen

e-mail: chong@ruc.edu.cn

B. Pei · T. Zhao · H. Chen

School of Information, Renmin University of China, Beijing, People's Republic of China

B. Pei

New Star Research Institute of Applied Tech, Hefei, People's Republic of China

Keywords Fuzzy associative classifier · Probabilistic numerical data · Data mining

1 Introduction

Classification is an important and well-researched data mining task, which is widely used in numerous real-world applications, such as customer evaluation and fraud detection. Generally speaking, classification is a supervised learning process of identifying to which of a set of classes a new instance belongs, on the basis of a training set of data containing instances whose class is given in advance. People have first proposed a large number of classification algorithms on the dataset with precise data, among which associative classifier approach is one of the most used classification methods because it is relatively easy to understand for human beings and often outperforms decision tree learners on many classification problems.

Recently, the importance of uncertain data is growing quickly in many essential applications, such as environmental monitoring, mobile object tracking, privacy protection, and data integration. Thus, researches on classification for uncertain data, especially on probabilistic data, have attracted increasing attention in the data mining field. Considering that associative classifier is one of the most important and useful classification methods for precise data, it is urgent to apply it for uncertain data from theoretical and application aspects. In the real-world applications, a probabilistic dataset with numerical attributes, such as income, age and price, is widely used. For example, in a habitat-monitoring system, the data measured from sensors like temperature and humidity are inherently uncertain due to the factory accuracy, the environmental erosion, its reduced power consumption, etc. Therefore, when we get a measured value “20” from a temperature sensor, we may not be sure it is true and can associate a probability, say, 20 %, with this value showing the confidence that this measured value is credible and true. In this case, the sensor value can be represented by a probabilistic numerical item (20, 20 %).

In this paper, we propose an approach to build an associative classifier for a *probabilistic numerical dataset* (PND). Firstly, by transforming the original PND to a probabilistic dataset with fuzzy sets, we introduce new definitions of support and confidence suitable for a dataset with both fuzziness and randomness. Based on these new measures, we then develop an Apriori-based algorithm to mine fuzzy association rules (FARs) from a PND. Secondly, due to the uncertainty of the data, an instance may be partially covered by a fuzzy classification rule. So we discuss a new method that can identify the weight of a probabilistic instance covered by a fuzzy association rule. We also introduce redundant pruning fuzzy rules method and instance coverage strategy to build fuzzy associative classifier in order to reduce the rules in the classifier while improving the accuracy and efficiency of classification. We lastly redefine the fuzzy multiple rules classification for new

instances. To the best of our knowledge, this is the first paper addressing the problem of associative classification for probabilistic numerical data.

Prior work. Recently, there have been researches on classification for uncertain data. These existing works all intend to extend classical classification algorithms, which focus on precise data, to probabilistic data. Qin et al. [2] introduced a rule-based algorithm to handle the problem of classifying uncertain data. Based on the traditional decision tree algorithm, the authors later proposed uRule algorithm to classify uncertain data by using new probabilistic information gain [3, 4]. Their method is suitable for uncertain categorical data and uncertain numerical data. However, their work is different from ours, because uncertain numerical data model used in [2–4] is denoted by a numerical interval and the probability distribution is defined in this interval, while probabilistic numerical data in our work is represented by a numerical value associated with a probability indicating the confidence that this value is true and credible. Qin et al. [5] extended classical Bayesian classification algorithm for classifying uncertain categorical and numerical data. The uncertain data model in this paper is the same as their previous work [2–4]. Ref. [6] extended classical decision tree building algorithms to handle uncertain data presented by multiple values forming a probability distribution function (PDF). Ref. [1] introduced uCBA algorithm for classifying only uncertain categorical data based on associative classifier algorithm. Gao et al. [7] presented uHarmony, which is based on traditional algorithm HARMONY, to solve the problem of classifying uncertain categorical data by mining discriminative patterns directly from uncertain data as classification features/rules and then helping train either SVM or rule-based classifier. However, there has been little research on associative classifier for probabilistic numerical data so far.

2 Problem Statement

In this section, we give the probabilistic numerical database model used in this paper. We then briefly introduce the framework of fuzzy associative classifier for probabilistic numerical data. In this paper, following studies in [2–4], we only consider probabilistic numerical attributes and assume the class label is certain.

2.1 The Probabilistic Numeric Data Model

When the value of a numerical attribute is uncertain, the attribute is called a *probabilistic numerical attribute*. Let $U = \{I_1, I_2, \dots, I_m\}$ be a set of numerical data type attributes. $I_j = \{i_{j1}, i_{j2}, \dots, i_{j1}, \dots, I_{jn}\}$, $1 \leq j \leq m, 1 \leq l \leq |DB|$, where i_{jl} represents the value of item in l th row and j th column. In the paper, fuzzy data are expressed by linguistic terms, that is, low, middle, and young. Suppose

fuzzy sets $F_j = \{F_{j1}, \dots, F_{jk}, \dots, F_{jK}\}$ are associated with attribute I_j , $1 \leq j \leq m$, $1 \leq k \leq K$.

A *probabilistic numerical item* (n -item) is defined as v_{jl} , p_{jl} , where v_{jl} represents the numerical value of item v_{jl} , and p_{jl} represents the existential probability associated with this value and thus should be in interval $[0, 1]$. For example, the value “(20, 20 %)” is an n -item measured by a sensor. This pair reflects that the measured value 20 is credible with a probability of 20 %, and meanwhile, the measured value is not 20 with a probability of 80 %.

Based on the notion of n -item, a PND is a dataset $D = \{t_1, \dots, t_l, \dots, t_n, \text{class}\}$, where $t_l = \{(v_{1l}, p_{1l}), \dots, (v_{jl}, p_{jl}), \dots, (v_{ml}, p_{ml})\}$ and (v_{jl}, p_{jl}) ($1 \leq j \leq m$, $1 \leq l \leq n$) is a n -item; class is the class attribute, which is known for certain.

Another kind of item used in the paper is *uncertain item*. An *uncertain item* (u -item) is denoted as $(f/F, p)$, where f is the membership degree for fuzzy set F , and p is the probability of (f/F) . Obviously, a u -item combines fuzzy uncertainty with random uncertainty, and it indicates the probability of fuzzy event (f/F) is p .

2.2 Fuzzy Association Rules from a PND

The association rules mined from a PND are FARs with the form of $(X, A) \Rightarrow (Y, B)$, where $X = \{x_1, x_2, \dots, x_p\}$, $A = \{f_1, f_2, \dots, f_p\}$, $Y = \{y_1, y_2, \dots, y_q\}$, $B = \{g_1, g_2, \dots, g_q\}$, $f_i \in \{\text{fuzzy sets related to attribute } x_i\}$, $g_j \in \{\text{fuzzy sets related to attribute } y_j\}$. A and B contain the linguistic-valued fuzzy sets associated with the corresponding attributes in X and Y .

The reason why we mine FARs from a PND is that there are n -items in this kind of dataset. Because the former part of an n -item is numerical data, “crisp partitioning” of such data may lead to undesirable boundary problems. Fuzzy set theory is thus recognized suitable to deal with the “sharp boundary” problem by providing a flexible remedy. By allowing of fuzzy sets in association rules mining from a PND, our proposed fuzzy associative classifier has better understandability in terms of knowledge representation and the smooth boundaries while keeping the satisfactory accuracy.

2.3 The Framework of our Proposed Algorithm

Before we give the details of our algorithm, we first give the framework of the algorithm in Fig. 1. The framework of our algorithm is similar to CBA [8] and CMAR [9], which are classical associative classifier algorithms for precise data.

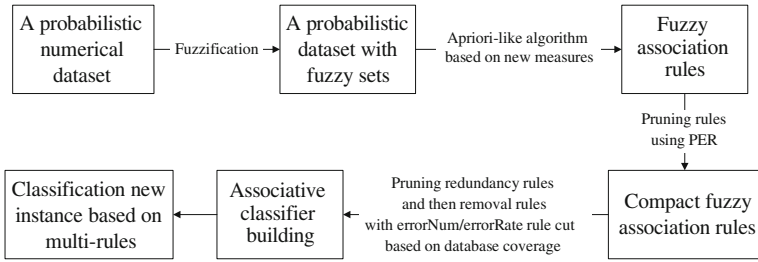


Fig. 1 The framework of our proposed algorithm

3 Mining Fuzzy Association Rules from a PND

It is well known that CBA and its modifications have been designed for precise data, but are not suitable for probabilistic numeric data any more since uncertainty is involved. In this paper, we introduce a new Apriori-like algorithm to discover FARs. We first convert a PND to a probabilistic database with fuzzy sets (i.e., mapping numerical data to its corresponding fuzzy sets). Next, we propose a novel metric to measure the *information content* (IC) of a *u*-item with coexistence of fuzzy and random uncertainty. Based on this measure, new support and confidence measures for the probabilistic database with fuzzy sets are introduced. We then give an Apriori-like algorithm to mine FARs.

3.1 New Measures

Now, we define support and confidence degrees for FARs in a PND. In a probabilistic numerical database, because of the uncertainty of *n*-items, a probabilistic transaction may partially support a rule, so we first propose a novel metric to measure the certain information of an item with both fuzzy and random uncertainties.

Considering *Shannon Information Entropy* is a well-defined and well-used uncertainty and information measure, we introduce it and redefine an entropy-like function to measure the uncertainty and information in such database.

Given an *n*-item $A = (\mu(A)|F)p(A)$, its *necessity* and *possibility measure* can be defined as

$$\Pi(A) = \max(\mu(A), p(A)) \tag{1}$$

$$N(A) = \min(\mu(A), p(A)) \tag{2}$$

where $\mu(A)$ is the membership degree for fuzzy set *F*, and $p(A)$ is the probability of the fuzzy event. In *possibility theory* [10], necessity measure $N(A)$ means the smallest possibility that item *A* would happen, and possibility measure $\Pi(A)$ means the largest possibility that item *A* would happen.

We then define a novel Shannon-like entropy metric to measure the certain IC that a u-item has. Given an n-item NI, its IC can be defined as follows:

$$IC(NI) = \frac{1}{c} (N(NI)En(N(NI)) + \Pi(NI)En(\Pi(NI))) \tag{3}$$

where

$$En(N(NI)) = \begin{cases} \frac{1}{2} Entropy(N(NI)), & N(A) \leq 0.5 \\ 1 - \frac{1}{2} Entropy(N(NI)), & \text{others} \end{cases}$$

$$En(\Pi(NI)) = \begin{cases} \frac{1}{2} Entropy(\Pi(NI)), & \Pi(A) \leq 0.5 \\ 1 - \frac{1}{2} Entropy(\Pi(NI)), & \text{others} \end{cases}$$

$$Entro(N(NI)) = -(N(NI) \log_2 N(NI) + (1 - N(NI)) \log(1 - N(NI)))$$

$$Entro(\Pi(NI)) = -(\Pi(NI) \log_2 \Pi(NI) + (1 - \Pi(NI)) \log(1 - \Pi(NI)))$$

$$c = N(A) + \Pi(A) + N(\bar{A}) + \Pi(\bar{A})$$

We can prove that $N(x)$ and $\Pi(x)$ increase, IC value increases, too. That is to say, IC of an n-item is monotone increasing, which is suitable for definition of support measure in association rule mining problems, because support measure is in fact a map from universe to interval $[0,1]$ satisfying increasing monotony. The proof is omitted here for lack of space.

Now, we give *support* and *confidence* based on IC measure. Given an itemset $UI = \{(I_1 F_{1i}), \dots, (I_p : F_{pj})\}$, the *support* measure of UI in database D is defined by the formula

$$Sup_D(UI) = \left(\sum_{i=1}^n IC_{t_i}(I_1.F) \otimes IC_{t_i}(I_2.F) \otimes \dots \otimes IC_{t_i}(I_p.F) \right) / n \tag{4}$$

where n is the total number of transactions in database D . $IC_{t_i}(I_j.F_j)$ ($1 \leq j \leq p$) is the IC measure of fuzzy set F associated with attribute I_j in t_i . \otimes is a T -norm.

Given a FAR $R = (X, A) \Rightarrow (Y, B)$, its support and confidence degrees are as follows:

$$Support(R) = Sup_D((X, A) \cup (Y, B)) \tag{5}$$

$$Confidence(R) = Sup_D((X, A) \cup (Y, B)) / Sup_D(X, A) \tag{6}$$

3.2 Pruning Rules Using Pessimistic Error Rate

Normally, many strong FARs may be discovered after mining from a PND. In fact, a large number of insignificant FARs make no contributions or even do harm to classification accuracy by introducing noise and may mislead from taking any role in the prediction process of test data objects. The removal of such rules makes the

classification process more accurate and efficient. In this paper, we also utilize pessimistic error rate (PER)-based pruning method presented in C4.5 and See5 to prune non-interesting rules. The error rate of rule R is formalized as:

$$q(R) = \frac{\left(r + \frac{z^2}{2n} + z\sqrt{\frac{r}{n} - \frac{r^2}{n} + \frac{z^2}{4n}} \right)}{\left(1 + \frac{z^2}{n} \right)} \tag{7}$$

where r is the observed error rate of rule R , that is, $r = 1.0 - \text{confidence}(R)$; n is the support of rule R , that is, $n = \text{Support}(R)$; $z = \Phi^{-1}(c)$, c is the confidence level given by the user.

The FAR-generating algorithm introduced here consists of two steps. First, a PND is converted to a probabilistic database with fuzzy sets using membership functions obtained by fuzzy clustering methods on numerical data in advance. Next, FARs are discovered from such database by Apriori-like algorithm based on new support and confidence measures. In this step, PER-based pruning method is used to prune some non-interesting rules. The algorithm is similar to Apriori, so it is omitted here for the lack of space.

4 Building Fuzzy Associative Classifier

After we discover FARs from the PND, there are still huge number of fuzzy rules that are generated, even though PER-based pruning method is used, which make no contributions to or even do harm to classification. In this section, we illustrate how to prune redundant fuzzy rules and then build the final classifier based on database coverage strategy.

4.1 Redundant Fuzzy Rules Pruning

Though fuzzy classification rules can be discovered using extended Apriori-like association rule mining techniques directly, the whole set of rules might be poor in quality. First, the number of rules may be too large to easily construct a classifier. More seriously, from the viewpoint of classification, there may exist conflicting rules and redundant rules. For example, $F \Rightarrow C1$ and $FF' \Rightarrow C1$ with $\text{Confidence}(F \Rightarrow C1) > \text{Confidence}(FF' \Rightarrow C1)$, then FF' is a redundant rule. The redundancy will result in some rules useless for classification.

Definition 4.1 Given two fuzzy classification rules R_1 and R_2 , R_1 is said to have **higher rank** than R_2 , denoted as $R_1 > R_2$, if and only if

- (1) $\text{Confidence}(R_1) > \text{Confidence}(R_2)$; or
- (2) $\text{Confidence}(R_1) = \text{Confidence}(R_2)$, but $\text{Support}(R_1) > \text{Support}(R_2)$; or

- (3) $\text{Confidence}(R_1) = \text{Confidence}(R_2)$, and $\text{Support}(R_1) = \text{Support}(R_2)$, but R_1 has fewer attribute values in its left-hand side than R_2 does.

Definition 4.2 A fuzzy classification rule R_1 is said a **general rule** w.r.t. fuzzy classification rule R_2 , denoted as $R_1 \succ R_2$, if and only if the antecedent part of R_1 is a subset of R_2 , and $R_1 > R_2$.

The purpose of defining *general rule* is to identify redundant rules. Given $R_1 \succ R_2$, if the consequents of them are the same, then R_2 is redundant. That is to say, R_2 should not be contained in the classifier once it is inferior to another rule. Thus, only general rules are left in the fuzzy rules.

4.2 Database Coverage Pruning Strategy

Even though redundant rules are removed, we still need to remove non-interesting fuzzy rules and select a subset of high-quality rules for classification, which can further improve the accuracy of the classifier.

Following classical CMAR algorithm, we propose a pruning rules strategy based on database coverage, but redesign the strategy in order to make it appropriate for fuzzy rules.

The key idea behind the database coverage strategy is that in classifier builder algorithm, a rule is selected if it can correctly classify an instance in the training dataset, instead of removing one training instance immediately after it is covered by one selected rule like CBA does; we let it stay there until its covered weight reached a user-defined threshold δ , which ensures that each training instance is covered by at least δ rules. This allows us to select more classification rules. Thus, when classifying a new data object, it may have more rules to consult and may have better chances to be accurately predicted.

However, the difficulty is that, in fuzzy-probabilistic dataset, due to the uncertainty of the data, an instance may be partially covered by a fuzzy classification rule. So the weight of the training instance will not increase by one when it is covered by a fuzzy classification rule, like CMAR does in certain dataset. In this paper, we define the weight as follows:

Given a fuzzy classification rule CU and a training instance t_i , suppose antecedent of CU and t_i share the same attributes and the same fuzzy set on these attributes: $\{(I_1 : F_{1i}), (I_2 : F_{2j}), \dots, (I_p : F_{pk})\}$, and if CU correctly satisfies t_i , then we define the coverage weight of the instance t_i covered by the rule CU as the following formula:

$$\text{weight}(t_i, \text{CU}) = \text{IC}_{t_i}(I_1.F_{1i}) \otimes \dots \otimes \text{IC}_{t_i}(I_p.F_{pk}) \tag{8}$$

where $\text{IC}_{t_i}(I_j.F_j)$ ($1 \leq j \leq p$) is the IC of fuzzy set F associated with attribute I_j in t_i , and \otimes is a T -norm. Thus, we prune an instance in the training set until the coverage weight of this instance reaches a user-defined threshold δ .

Algorithm 1. Classifier Builder Algorithm

Input: a training probabilistic dataset D ; a set of fuzzy classification rules generated after removing redundant fuzzy rules R ;

Output: the final classifier C .

Method:

```

Initialize  $C = \phi$ ,  $\text{totalWeight}[i]=0, i \in [1, |D|]$ ; // the total coverage weight of the  $i^{\text{th}}$  instance
for each fuzzy classification rule in sequence {
  for each training instance {
    if  $\text{totalWeight}[d.\text{id}] \leq \delta$  &&  $d$  satisfies the conditions of  $r$  && it is correctly
classifies  $d$  {
       $\text{totalWeight}[d.\text{id}] = \text{totalWeight}[d.\text{id}] + \text{weight}(d, r)$ ;
       $C = C \cup r$ ;
      Select a default class of the current  $C$  for the remaining training dataset;
      //two methods to compute the error rate of the current classifier
      Compute the total number of errors of the current  $C$ ;
    } end if
  } }
Find the first fuzzy rule  $k$  in  $C$  with the lowest error rate and drop all the rules after  $k$  in  $C$ ;
Add the default class associated with  $k$  to end of  $C$ ;
return  $C$ ;
```

Fig. 2 Pruning fuzzy rules to build classifier

On the other hand, each time when a rule is selected in a temporary classifier, the error rate of the current temporary classifier is calculated. In this paper, we have adopted two methods to compute the error rate of a classifier:

1. *errNum* method: Apply the classifier C to the training set of the dataset and count the number of instances that are misclassified using the classifier. That is the error rate of C .
2. *errRate* method: For the classifier C , the error rate of C is defined by the formula:

$$ER = \text{num_misclassify} / \text{num_misclassify}, \quad (9)$$

where num_corclassify is the number of the instances correctly classified by C , while num_misclassify is the number of the instances incorrectly classified by C .

After we cut these non-interesting fuzzy rules, the rule with the lowest error rate, a fuzzy associative classifier is finally built (Fig. 2).

5 Classification Based on Multi-rules

After the fuzzy associative classifier has been built, as discussed in Sect. 4, we are ready to classify new instances. In this section, how to classify new instances with the selected fuzzy classification rules in the classifier is discussed.

We first define a match measure (namely CD) between a fuzzy classification rule and a new instance.

Definition 5.1 Given a fuzzy rule $r: F \Rightarrow C$ and a new instance t , the confidence degree of classifying t with r is as follows:

$$CD(t, r) = \text{Confidence}(r) * \text{weight}(t, r), \quad (10)$$

where $\text{confidence}(r)$ is the degree of confidence for rule r , and $\text{weight}(t, r)$ is the match weight of instance t covered by rule r .

5.1 Classification Based on One Rule

Classification based on one rule is processed as follows: Given a new instance t , the rule with the highest DF is used to classify t . Notably, for crisp rules, the value of $\text{weight}(t, r)$ either is 1 where the rule r covers t or is 0 where r does not cover t . Thus, the rule r that covers t with the highest confidence will be chosen to classify t , which is the same with CBA and GARC classifiers. This method is so called *single rule classification*.

However, this direct approach ignores uncertain information in the instance and hence may decrease the prediction accuracy.

5.2 Classification Based on Multi-Rules

The key idea of classification based on multi-rules is that, given a new case t , a set of related fuzzy classification rules (i.e., those rules that match t) in the classifier is selected to classify with a compound measure to determine which label will be given to t .

There are two cases with the set of rules:

If all the selected related rules have the same label, the algorithm simply assigns the same class label to t ;

If the selected related rules have different labels, the algorithm first divides the rules into several groups according to the class label. Those rules that have the same label are in the same group. Next, the algorithm uses an index to measure *the strength of the class group*.

While CMAR performs classification based on a weighted χ^2 analysis by using multiple strong associative rules, this does not work well in the context of fuzzy-probabilistic dataset. We follow the idea of classification based on multi-rules and

Algorithm 2. Classification a new instance based on multi-rules

Input: Fuzzy classifier R generated by algorithm 3; a new instance t that needs classify.

Output: c: the classification predicted for instance t.

Method:

```

Initialize SW( $r_{c_p}$ )=0; //p ∈ [1,i];
for each fuzzy classification rule in sequence {
    if r satisfies the conditions of instance t
        case the class label of rule r is //Suppose the //class labels can be divided to i groups
//according to rules that satisfy instance t
        c1: SW( $r_{c_1}$ ) += CD(t,  $r_{c_1}$ );
        c2: SW( $r_{c_2}$ ) += CD(t,  $r_{c_2}$ );
        .....
        ci: SW( $r_{c_i}$ ) += CD(t,  $r_{c_i}$ );
        endcase
    endif
}
Predict the class label, c, of instance t with the label that has the maximum SW;
return c;

```

Fig. 3 Classifying a new instance algorithm

propose a simple but effective way to measure the compound effect of the group rule.

Given a group of fuzzy rules with the same class label CL: $R_{CL} = \{r_1, r_2, \dots, r_n\}$. The *strength weight* of these rules is defined by the following formula:

$$SW(R_{CL}) = \sum_{i=1}^{|D|} DF(t, r_i) \tag{11}$$

Given a new instance t, suppose the correlated rules with t have been divided into p groups according to the class labels: $\{CL_1, CL_2, \dots, CL_p\}$. Then the class label of t is computed as follows:

$$\begin{aligned} \text{ClassLabel}(t) &= CL_j, \\ \text{if } SW(CL_j) &= \max(SW(R_{CL_1}), SW(R_{CL_2}), \dots, SW(R_{CL_p})) \end{aligned} \tag{12}$$

Fig. 3 gives the algorithm for classifying a new instance based on multi-fuzzy rules.

6 Experimental Results

In this section, we will present the experiments of our proposed classification algorithm for a PND.

Table 1 Information of the datasets

Dataset	No. of instances	No. of Attributes	Class attribute type
Iris	150	5	Nominal
Wine	178	13	Nominal
Heart	255	13	Nominal
Vertebral	310	7	Nominal
Breast Cancer	699	10	Nominal
Blood	748	5	Nominal

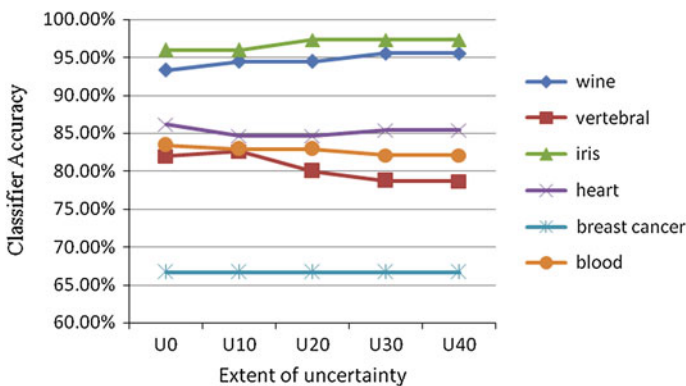
6.1 Experimental Setup

The experiments were conducted over 6 benchmark datasets from UCI repository datasets. The details of those datasets used can be found in Table 1. We executed the experiments on a PC with Intel Core 1.99 GHz and 992 MB main memory.

Following references [1–4], for each dataset, we select k attributes with maximum *information gain* values. We then add random uncertainty to the numerical values in these top K attributes into uncertain ones. When we introduce $M\%$ uncertainty to these top k attributes, this means that $M\%$ item value of this attribute will be added to a random probability, and the left $(1 - M)\%$ item value of this attributes should be added to 100% probability. Thus, the PND is denoted as TopKuM (Top K attributes with $M\%$ uncertainty).

6.2 Impact of the Level of Uncertainty

We first analyze the impact of the level of uncertainty in a PND on the accuracy and the classifier construction time. Figure 4 shows the accuracy of our classifier that remains relatively stable. From the figure, the classification accuracy of our

**Fig. 4** The accuracy of the datasets in different levels of uncertainty

method is quite satisfactory. Figure 5 shows the classifier construction time remains relatively stable too. It is because that the dominant of classifier construction time is the size of dataset, which can be easily seen from the Fig. 5. Because the construction time mainly contains the fuzzification time, mining time, and a fraction of pruning time, however, fuzzification time and mining time are closely related to the size of datasets, so the construction time remains stable even though the level of uncertainty varies.

6.3 Impact of δ on the Number of Classification Errors

As discussed earlier, the parameter δ controls the number of rules contributed to an instance. The larger the δ is, the more rules we select, thus reducing the number of classification errors and improving the accuracy. Figure 6 confirms this.

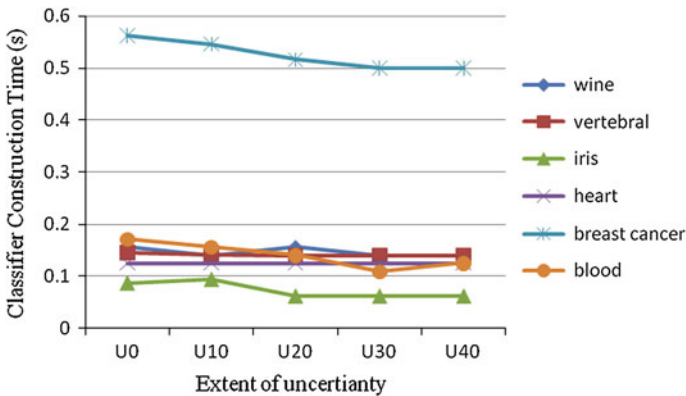


Fig. 5 The classifier construction time in different levels of uncertainty

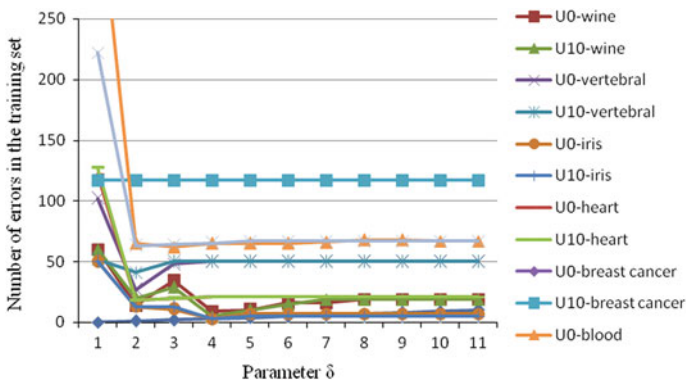


Fig. 6 The impact of δ on the number of errors in the training set

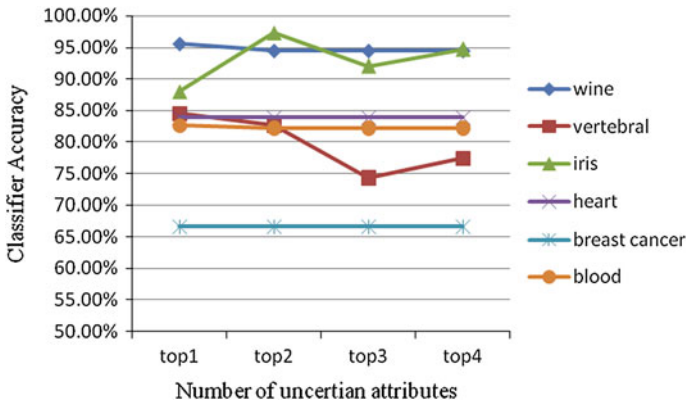


Fig. 7 The impact of number of uncertain attributes on classifier accuracy

In addition, we can see that when the δ exceeds 4, the performance of our classifier over the 6 datasets with uncertainty tends to be stable. Therefore, we set parameter δ to 4 in all of the other experiments in this paper.

6.4 Impact of the Number of Uncertain Attributes

These experiments study the number of uncertain attributes' effect to the accuracy of our classifier. As shown in Fig. 7, the overall tendency stays stable. With increasing in uncertainty attributes, the IC of every attribute would be decreasing. However, the values of support and confidence are not changing; this factor would affect the number of the original rules mined from a PND. At last, the accuracy of our classifier would be waved to some extent.

6.5 Comparative Study Between Single Rule Classification and Multiple Rules Classification

In this experiment, we evaluate the performance of our algorithm on different levels of uncertain dataset (U0–U70), while using the single rule classification method and the multiple rules classification method, which is described in Sect. 5. In Table 2, column *Single* represents accuracy of single rule classification, and column *Multiple* represents accuracy of multiple rules classification. We computed the accuracy for every datasets in different levels of uncertainty.

We only show a part of the results for the lack of space. As shown in Table 2, with increasing in uncertain level, the accuracy of our approach in both single rule classification and multiple rules classification fluctuated to some extent. Overall,

Table 2 The classifier accuracy of single versus multiple rules classification

Dataset	U20		U30		U40	
	Single (%)	Multiple (%)	Single (%)	Multiple (%)	Single (%)	Multiple (%)
Wine	Single	Multiple	Single	Multiple	Single	Multiple
Vertebral	64.45	94.44	65.56	95.56	62.22	95.56
Iris	68.39	80	69.03	78.71	67.74	78.71
Heart	66.67	97.33	66.67	97.33	66.67	97.33
Breast cancer	83.85	84.62	83.08	85.38	83.08	85.38

Table 3 The classifier accuracy of errNumber versus errRate rule cut method

Dataset	U20		U30		U40	
	errNum (%)	errRate (%)	errNum (%)	errRate (%)	errNum (%)	errRate (%)
Wine	78.89	94.44	78.89	95.56	83.33	95.56
Vertebral	67.74	80	67.74	78.71	67.74	78.71
Iris	92	97.33	93.33	97.33	94.67	97.33
Heart	83.85	84.62	83.85	85.38	83.85	85.38
Breast cancer	64.1	66.67	64.1	66.67	64.1	66.67
Blood	82.09	82.89	82.09	82.09	82.09	82.09

on all the uncertain datasets, the accuracy of the classifier using *Multiple* rules for classification exceeds that of the classifier using *Single* rule.

6.6 Comparative Study Between the errNum Rule Cut and errRate Rule Cut

In Sect. 4.2, we adopted a traditional rule cut method, *errNumber*, and a new rule cut method, *errRate*, to measure whether the rule could be added to final rule set or not. So, in this experiment, we study the effect of these two methods on the accuracy of classifier.

We can see from Table 3 that the accuracy of the classifier by using our proposed *errRate* method is higher than that of using *errNum* method. So we adopt *errRate* cut method in the database coverage when building the final classifier in all of the other experiments in this paper.

7 Conclusions

This paper proposes a generic framework of fuzzy associative classifier for PND. Based on new support and confidence measures, FARs are mined from a probabilistic dataset with fuzzy sets, which is converted by the original PND. We then

introduce the redundant fuzzy rules pruning method and database coverage strategy to build an associative classifier suitable for PNDs. We also redefine multiple fuzzy rules to classify new instances. Our experimental results demonstrate that the proposed algorithm has satisfactory performance on different kinds of probabilistic numerical data.

Acknowledgments This research was supported by the National Basic Research Program of China (973program) (2012CB316205), the National Natural Science Foundation of China (61070056, 61033010, 61202114), the HGJ Important National Science & Tech Specific Projects of China (2010ZX01042-001-002-002), and the Fundamental Research Funds of Renmin University of China (12XNLF07).

References

1. Qin XJ, Zhang Y, Li X, Wang Y (2010) Associative classifier for uncertain data. In: The 11th international conference on web-age information management (WAIM), pp 692–703
2. Qin B, Xia Y, Prbahakar S, Tu Y (2009) A Rule-based classification algorithm for uncertain data. In: IEEE international conference on data engineering (ICDE), pp 1633–1640
3. Qin B, Xia Y, Li F (2009) DTU: a decision tree for uncertain data. In: The Pacific-Asia conference on knowledge discovery and data mining (PAKDD), pp 4–15
4. Qin B, Xia Y, Prabhakar S (2011) Rule induction for uncertain data. *Knowl Inf Syst* 29:103–130
5. Qin B, Xia Y, Li F (2010) A Bayesian classifier for uncertain data. In: ACM symposium on applied computing (SAC), pp 1010–1014
6. Tsang S et al (2011) Decision trees for uncertain data. *IEEE Trans Knowl Data Eng* 23(1):64–78
7. Gao CC, Wang JY (2010) Direct mining of discriminative patterns for classifying uncertain data. In: ACM SIGKDD international conference on knowledge discovery and data mining (KDD), pp 861–870
8. Liu B, Hsu W, Ma YM (1998) Integrating classification and association rule mining. In: ACM SIGKDD international conference on knowledge discovery and data mining (KDD) 1998, pp 80–86
9. Li WM, Han JW, Pei J (2001) CMAR: accurate and efficient classification based on multiple class-association rules. In: IEEE international conference on data mining, pp 369–376
10. Dubois D, Prade H (1988) Possibility theory: an approach to computerized processing of uncertainty. Kluwer Academic/Plenum Publishers, New York

Adaptive Boosting for Enhanced Vortex Visualization

Li Zhang, Raghu Machiraju, David Thompson,
Anand Rangarajan and Xiangxu Meng

Abstract In this paper, we demonstrate the use of machine learning techniques to enhance the robustness of vortex visualization algorithms. We combine several local feature detection algorithms, which we term weak classifiers into a robust compound classifier using adaptive boosting or AdaBoost. This compound classifier combines the advantages of each individual local classifier. Our primary application area is vortex detection in fluid dynamics datasets. We demonstrate the efficacy of our approach by applying the compound classifier to a variety of fluid dynamics datasets.

1 Introduction

As computer power continues to increase, the complexity of simulations, in terms of both the physics modeled and the simulation size, also increases. Future exa-scale computing systems will generate increasingly larger simulation datasets [1]. Even now, data are being produced at a rate that far exceeds the ability of application scientists to analyze it. What is lacking are the tools needed to facilitate data analysis and visualization of the resulting massive quantities of data.

L. Zhang (✉) · X. Meng
Shandong University, Shandong, China
e-mail: lizhangchina@hotmail.com

R. Machiraju
The Ohio State University, Columbus, OH, USA

D. Thompson
Mississippi State University, Starkville, MS, USA

A. Rangarajan
University of Florida, Gainesville, FL, USA

One potential tool set, feature detection, is already an important data strategy for scientists who deal with terascale and petascale data. Fundamentally, feature detection operates by reducing the amount of data that needs to be analyzed to a set of feature descriptors or a feature catalog. There are two distinct paradigms that can be employed to identify feature of interest in a dataset [2]: local and global. The local approach, or point classification, operates on a small neighborhood of the data and performs a binary classification as to whether a specific data point belongs to a feature (e.g., shocks in flow data). The collection of identified data points can then be aggregated to form the feature. In contrast, a global approach identifies a feature in its entirety by an aggregate classification strategy and many of them require information from nonlocal regions of the dataset (e.g., streamlines in flow data). For some feature types, the global approach can be more discriminating in terms of identifying features of interest.

In many fluid dynamics applications, vortices are the feature of interest. There are many algorithms that have been created to identify vortices. Unfortunately, they may encounter situations in which the detector incorrectly indicates the presence of feature (false positive) or fails to locate an existing feature (false negative) [3–5]. These occur because the detector is not based on a formal, rigorous definition.

We conceptualize this as the problem of robustness in feature detection and attempt to address it via boosting methodologies. We aim to judiciously combine different detection algorithms by using boosting methodologies. The local feature is what we concern about. Our proposed approach is to combine several local feature detection algorithms into a single compound classifier using adaptive boosting (AdaBoost) [6] that results in validated and robust feature detection. Ideally, the compound classifier would combine the best of all local classifiers as they respond to the underlying physical signal. Presumably, the classifiers created by AdaBoost would converge asymptotically to the ideal classifier.

Our paper is structured as follows. First, we summarize the vortex detection methods that we propose to employ as candidate weak classifiers and discuss some of the issues associated with vortex definition in Sect. 2. Then, in Sect. 3, we discuss the machine learning techniques that we use in the algorithm and describe our algorithm. We then demonstrate the resulting algorithm, provide concluding remarks, and suggest potentially fruitful directions for further research.

2 Related Work

2.1 Vortex Detection

There is no consensus on a formal definition of vortex in the literature. The most intuitive description of a vortex is based on the notion of swirling fluid motion. Robinson [7] describes a vortex in terms of its instantaneous streamlines as:

A vortex exists when instantaneous streamlines mapped onto a plane normal to the vortex core exhibit a roughly circular or spiral pattern, when viewed from a reference frame moving with the center of the vortex.

However, appealing this description is self-referential. That is, to find a vortex, you must first know where it is and how fast and what direction it is moving. Despite the lack of a formal definition, various vortex detection algorithms have been developed that have demonstrated success identifying vortices in computational datasets [8, 9]. Each of the various detection algorithms has an implicit definition of a vortex. The success, or lack thereof, of each method depends on how well its particular vortex definition matches the flow field in a given region.

The most intuitive approach, extracting streamlines or pathlines, is inherently a global method. Among the most sophisticated vortex detection algorithms are the one described by Haller [10] that provides an objective definition of a vortex based on the stability of fluid trajectories in unsteady, incompressible flows based on the M_Z criterion. It is too clearly global in nature because it requires trajectory computations. Similarly, the work of Garth et al. [11], which employs finite-time Lyapunov exponents (FTLE) [12] to characterize Lagrangian coherent structures, is also global in nature.

Field-type methods are good examples of local methods. In a field-type method, a scalar is defined at each point in the computational domain [13–19]. These scalar fields may be functions of the velocity, the velocity gradient, the pressure, or other field-related quantities, all of which are local quantities. Thus, these methods can be implemented using only a small neighborhood in the region surrounding the point under consideration. Local, field-type methods are ideal for use with exascale data because only local data are needed.

Topology-based methods [20–25] exploit the fact that there is a critical point in the velocity field at the vortex core in the plane containing the swirling motion. By their very nature, these methods provide a description of a vortex in terms of its core line or core region [20]. Each of these methods is also local in nature in that only a local neighborhood is needed for its implementation. These methods could also be implemented using a binary flag to mark nodes of the cells in which the vortex resides. However, the utility of these, and other critical point-based approaches, is somewhat limited because they are not Galilean invariant; therefore, they cannot be applied to translating vortices.

Unfortunately, as reported in the literature [3–5], none of these vortex detection schemes is foolproof. That is, they all have notable failures—both positives and false negatives. One approach that is employed to attempt to mitigate this drawback is to use combinations of detection methods [26–29] to make use of the favorable characteristics of each technique. For example, Burger et al. [26] express local binary feature detectors as fuzzy sets that can be combined using linking and brushing in an interactive visual framework. The primary issue associated with these approaches is the method of combining the results of the different detection algorithms.

2.2 Detection Algorithm

Since our overall objective is to facilitate exploration of large-scale fluid dynamics datasets, we want to employ local, field-type methods in the vortex detection process. Additionally, we are interested only in those methods that are Galilean invariant. Of the field-type methods, this eliminates both the normalized helicity [17] and the swirl parameter [13].

We now summarize the candidate methods that we propose to investigate as potential components of an enhanced vortex detection strategy. In the discussion that follows, we make reference to the rate of strain tensor S and the rate of rotation tensor Ω , which are defined in terms of the velocity gradient tensor J as:

$$S = \frac{J + J^T}{2}, \Omega = \frac{J - J^T}{2} \tag{1}$$

In some cases, e.g., two-dimensional steady flow, several of these methods reduce to the same approach. This is not the case, however, in more complex three-dimensional flows. The approaches discussed below will be implemented as a binary classifier in which a scalar field Σ_i is set to unity or zero based on whether or not the i -criterion is satisfied.

The Q-criterion [15] is based on the observation that, in regions where $Q = \frac{\|\Omega\|^2 - \|S\|^2}{2} > 0$, rotation exceeds strain and, in conjunction with a pressure minimum, indicates the presence of a vortex.

Δ -criterion [14] assumes that a vortex occurs in a region in which the eigenvalues of J include a complex conjugate pair. Here, $\Delta = \left(\frac{R}{3}\right)^3 + \left(\frac{\det J}{2}\right)^2 > 0$ indicates the presence of complex eigenvalues, where

$$R = \frac{\Omega_{ij}\Omega_{ji} + S_{ij}S_{ji}}{2} \tag{2}$$

However, relatively large regions of the flow can satisfy this criterion.

The λ_2 -method [16] defines a vortex to be connected region in which $\lambda_2 < 0$, where $\lambda_1 \leq \lambda_2 \leq \lambda_3$ are the eigenvalues of $S^2 + \Omega^2$. This corresponds to a region in which a rotation-induced pressure minimum occurs.

Another vortex center identification algorithm has been proposed by Michard et al. [30]. Let P be a fixed point in the measurement domain. He defines the dimensionless scalar function Γ_1 at P as

$$\Gamma_1(P) = \frac{1}{S} \int_{M \in S} \frac{(PM \wedge U_M) \cdot z}{\|PM\| \cdot \|U_m\|} dS = \frac{1}{S} \int_S \sin(\theta_M) dS \tag{3}$$

where S is a two-dimensional area surrounding P , M lies in S and z is the unit vector normal to the measurement plane. U_M represents the angle between the velocity vector U_M and the radius vector PM . It can be shown that $|\Gamma_1|$ is unity at the location of vortex center. In the same paper, they propose a way to determine

vortex boundary by approximating Γ_2 which is a local function depending only on Ω and μ .

$$\Gamma_2(P) = \frac{1}{S} \int_{M \in S} \frac{[PM \wedge (U_M - \tilde{U}_P)] \cdot z}{\|PM\| \cdot \|U_M - \tilde{U}_P\|} dS \quad (4)$$

where $\tilde{U}_P = (1/S) \int U dS$. They identify the region with $|\Omega/\mu| \geq 1$ is the vortex where Ω is the rotation rate corresponding to the antisymmetric part of the velocity gradient ∇u at P and μ is the eigenvalue of the symmetric part of this tensor.

3 Machine Learning to Enhance Robustness

The described algorithm above uses kinematic or dynamical properties (vorticity or pressure) or derived quantities (λ_2 , Δ -criterion, or Q -criterion) to classify nodes that are located within vortices. However, because these techniques are not based on a formalized definition of a vortex and because they depend upon noisy gradient computations or, in some cases, thresholding, none of these algorithms, in isolation, provide the best or optimal characterization of the flow features. We will implement each of the feature identification algorithms as binary classifiers. All classifiers are deemed weak pointwise since they do not have the benefit of being able to incorporate information from a large subset of the data. Thus, it is natural to create a composite classifier as an ensemble of weaker classifiers. The composite classifier, by design, has better performance on a training set and to the extent that the composite classifier's performance generalizes to unseen data, and the performance on the new data is expected to exceed that of the pointwise weak classifiers.

AdaBoost [6], short for Adaptive Boosting, is a meta-algorithm that can be used in conjunction with other machine learning algorithms to improve their performance. AdaBoost assumes the existence of weak learners-classification methods that are simple but perform only slightly better than chance and then boost their performance by judiciously combining the weak learners into a strong classifier. As illustrated in Algorithm 1, AdaBoost calls a weak classifier repeatedly in a series of rounds $t = 1, \dots, T$. For each call, a distribution of weights D_t is updated that indicates the importance of examples in the dataset for the classification. In each round, the weights for each incorrectly classified example are increased (or alternatively, the weights for each correctly classified example are decreased), so that the new classifier focuses more on those examples (Fig. 1).

Here, we propose to use AdaBoost to develop a framework to create consistently better classifiers for exascale data derived from physical simulations. The advantage of this approach is that other factors and cues can be easily

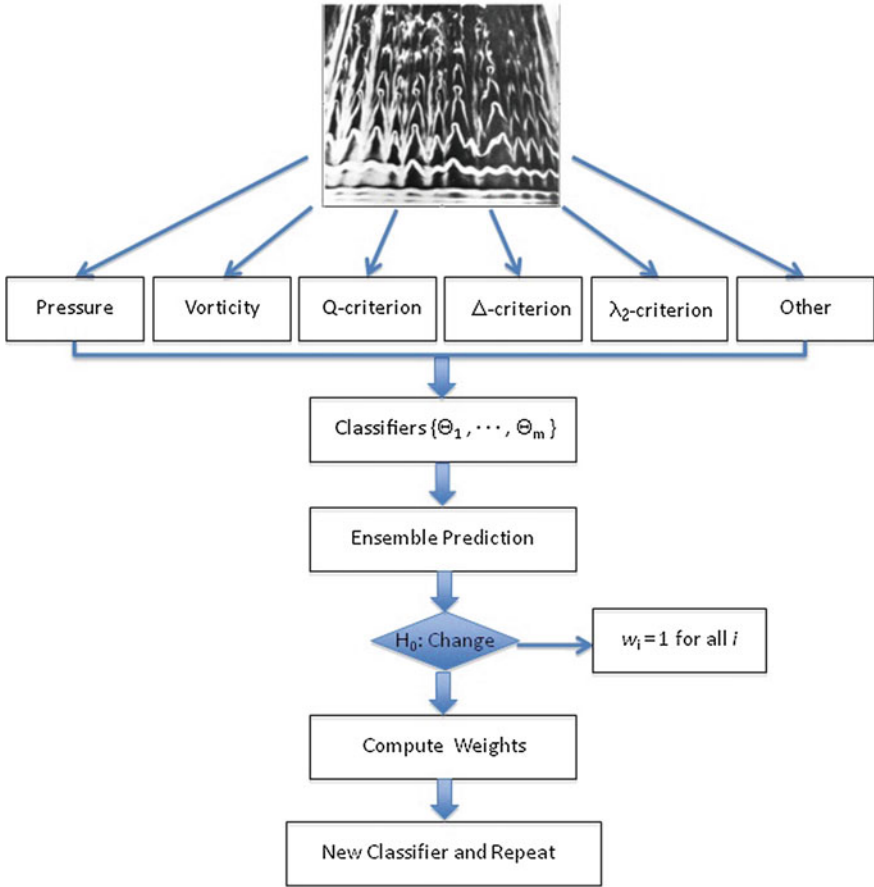


Fig. 1 An incoming stream is first classified by M weak classifiers manifest as various feature identification algorithms (followed by a thresholding operation). The stronger ensemble classifier realized as a linear combination of the weaker ones is then used to predict the membership of 4D points to specific flow features. If there is a dramatic change in the distribution as determined by a significance test, the weights are all reset to unity. Otherwise, they are adjusted at all instances of misclassification

incorporated. AdaBoost calls a weak classifier (or any of the feature identification methods) repeatedly in a series of rounds. For each call, a distribution is updated that indicates the importance of examples in the dataset for the classification. AdaBoost can be used in conjunction with other classifiers including the decision tree where the weights are continuously changed. Algorithm 1 provides a summary of the boosting process.

4 Experiments

In this section, we empirically validate our proposed algorithm by using experts to identify the regions of the domain that contain vortices area in fluid dynamics datasets.

Algorithm 1 AdaBoost algorithm

1. Given samples $(x_1, y_1), \dots, (x_n, y_n)$ where $y_i = 0, 1$ for negative and positive samples respectively.
2. Initialize weights $w_{1,i} = \frac{1}{2m}, \frac{1}{2l}$ for $y_i = 0, 1$ respectively, where m and l are the number of negatives and positive samples respectively.
3. For $t=1, \dots, T$:
 - (a) Normalize the weights,

$$w_{t,i} = \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}} \tag{5}$$

- (b) For each weak classifier h_j compute the error which is evaluated with respect to $w_t, \epsilon_j = \sum_i w_i ||h_j(x_i) - y_i||$.
- (c) Choose the classifier, h_t , with the lowest error ϵ_t .
- (d) Update the weights:

$$w_{t+1,i} = w_{t,i} \beta_t^{1-\epsilon_i} \tag{6}$$

where $\epsilon_i = 0$ if sample x_i is classified correctly by classifier h_t , $\epsilon_i = 1$ otherwise, and $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$.

4. The final strong classifier is:

$$h(x) = \begin{cases} 1 & \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where $\alpha_t = \log \frac{1}{\beta_t}$

4.1 Expert Labels

All the data must be prepared in advance of the training stage. These data include feature labels, weak hypothesis, and neighbor matrix. Here, the feature labels mean if these area are vortices or not (1 or 0). We rely on streamline visualization to identify regions containing vortices. The training samples are provided by domain experts aided by our visualization. The experts first define a block where could be circled labels easily, such as some relatively strong vortex with less turbulence. Then, expert chooses all positive samples with the cooperation of our interface.

For some datasets with same or similar condition, expert labels are provided once. For example, for a dataset with many time steps, it is necessary to expert to interactive with some of these time steps. It means these datasets share the same compound classifier. On the other hand, for completely different datasets, experts should provide labels for every single dataset.

4.2 Weak Classifiers

In our experiments, the weak classifiers are the λ_2 -method, the Q -criterion, the Δ -criterion, and the Γ_2 method. Since each weak classifier is sensitive to its given threshold, we use statistical methods to evaluate different thresholds to identify a suitable one for each of these weak classifiers. Among the various statistical methods, we use type I error, type II error, sensitivity, specificity, and accuracy in our experiments.

We rely on the receiver operating characteristic (ROC) curve, which is a graphical plot of the sensitivity, or true positive rate, versus false-positive rate, for a binary classifier system as its discrimination threshold is varied.

4.3 Results

We now apply our method on two different datasets: the tapered cylinder dataset [31] and delta wing dataset [32]. The tapered cylinder dataset is three-dimensional, incompressible, viscous flow around a cylinder that is transverse to the primary flow direction. Since the tapered cylinder dataset is time-varying dataset with 1000s of time steps, we trained a compound classifier in one time step. This compound classifier can be applied to other time steps because they have same condition. First, a domain expert marks vortex region in 2,000 samples in one single time step dataset, which defines the labels. We divide the samples into two parts: one part provides the training samples, and the second provides the testing samples. According to the training labels, we compute the best threshold for each weak classifier. The best threshold is the one that produces the best classification in the ROC space.

AdaBoost process gives us a final strong classifier:

$$h(x) = \begin{cases} 1 & 1.07117 * h_{\lambda_2} + 0 * h_{\Delta} + 123071 * h_Q + 06184 * h_{\Gamma_2} \geq 1.46013 \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

Table 1 shows a statistical characterization of the success of the weak classifiers and the compound classifier that includes accuracy (ACC), false-positive rate (FDR), specificity (SPC), positive predictive value (PPV), and negative predictive value (NPV). The compound classifier has a lower false-positive rate, a lower

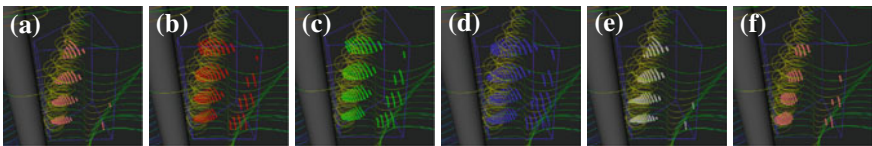
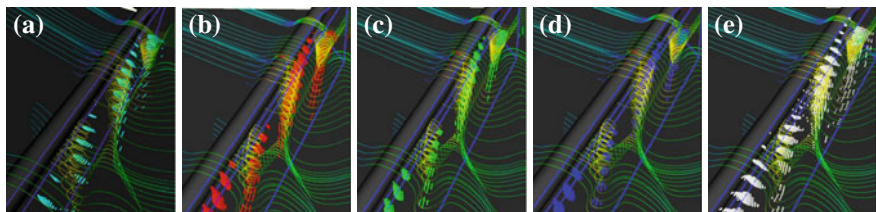


Fig. 2 Visualization results of different classifiers working on tapered cylinder dataset. **a** Compound classifier. **b** λ_2 . **c** Δ . **d** Q . **e** Γ_2 . **f** Expert Label

Table 1 Compare compound classifier to weak classifiers

	λ_2	Δ	Q	Γ_2	Compound classifier
ACC	0.5640	0.5580	0.5507	0.6567	0.8533
FDR	0.7919	0.7947	0.7996	0.7003	0.6653
NPV	0.9947	0.9960	0.9959	0.9922	0.9833
PPV	0.2081	0.2053	0.2004	0.2997	0.3347
SPC	0.6830	0.6739	0.6639	0.8116	0.8542

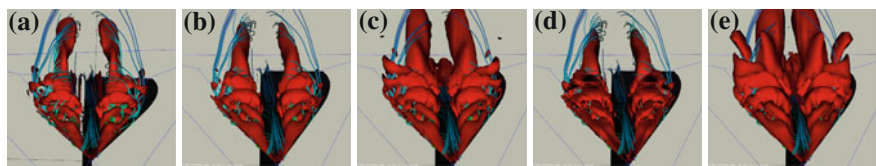
**Fig. 3** Visualization results of different classifiers. **a** Compound classifier. **b** λ_2 . **c** Δ . **d** Q. **e** Γ_2

NPV, a higher accuracy, a higher PPV, and a higher specificity. Figure 2 shows the visualization of comparison, where the dots indicate the vortex region. Then, we apply the compound classifier to another tapered cylinder dataset with different time step. As shown in Fig. 3, the compound classifier produces with a clearer vortex outline and less noise.

We also did an experiment on the delta wing dataset, which is defined on a $67 \times 209 \times 49$ curvilinear grid. Experts provided another 5,000 expert labels on a defined block of the delta wing. Here, we use all these 5,000 expert labels as training samples. We trained all the weak classifiers mentioned above to get a new compound classifier then applied it to the delta wing dataset. The compound classifier is:

$$h(x) = \begin{cases} 1 & 1.1126 * h_{\lambda_2} + 0.5352 * h_{\Delta} + 0.8642 * h_Q + 0 * h_{\Gamma_2} \geq 1.256 \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

Figure 4 shows isosurfaces of the compound classifier as well as the other classifiers. The compound classifier shows a cleaner, better defined vortex than the weak classifiers.

**Fig. 4** Visualization results of different classifiers working on delta wing dataset with isosurface. The red body is isosurface, and the blue line is streamline. **a** Compound classifier. **b** λ_2 . **c** Δ . **d** Q. **e** Γ_2

5 Conclusion

We have presented machine learning–based enhancement to vortex visualization in complex flow fields. This algorithm combines several different vortex detection algorithms, which we term weak classifiers, using a semi-supervised, adaptive boosting algorithm (AdaBoost). We compute a threshold value for each of the weak classifiers using the ROC space from the ideal prediction. Then, based on expert labeling, we computed a set of weights to be applied to each of the weak classifiers in order to produce a compound classifier. We used a well-known computational fluid dynamics dataset, the tapered cylinder, for training and testing the resulting compound classifier. Compared with the individual weak classifiers, the compound classifier shows a more accurate classification with lower false-positive rate, higher accuracy, and higher specificity. We also applied the composite classifier to a delta wing dataset with complex vortical structures.

Acknowledgments This work has been funded by the China National Natural Science Foundation(Grant No. U1035004 and 61003149).

References

1. ExaScale Software Study (2009) Software challenges in extreme scale systems. Technical report, DARPA
2. Thompson DS, Machiraju R, Jiang M, Nair J, Craciun G, Venkata S (2002) Physics-based feature mining for large data exploration. *IEEE Comput Sci Eng* 4(4):22–30
3. Chakraborty P, Balachandarand S, Adrian RJ (2005) On the relationships between local vortex identification schemes. *J Fluid Mech* 535:189–214
4. Haines R, Kenwright DN (1999) On the velocity gradient tensor and fluid feature extraction. In: *AIAA 14th Computational fluid dynamics conference*, paper 99-3288
5. Jiang M, Machiraju R, Thompson DS (2002) Geometric verification of swirling features in flow fields. *IEEE Vis* 02:307–314
6. Freund Y, Schapire R (1997) A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci* 55(1):119–139
7. Robinson SK (1991) Coherent motions in the turbulent boundary layer. *Ann Rev Fluid Mech* 23:601–639
8. Jiang M, Machiraju R, Thompson DS (2005) Detection and visualization of vortices. In: *Visualization handbook*. Academic Press, pp 287–301
9. Roth M (2000) Automatic extraction of vortex core lines and other line-type features for scientific visualization. PhD thesis, Swiss Federal Institute of Technology Zurich
10. Haller G (2005) An objective definition of a vortex. *J Fluid Mech* 525:126
11. Garth C, Gerhardt F, Tricoche X (2007) Efficient computation and visualization of coherent structures in fluid flow applications. *IEEE Trans Vis Comput Graph* 13(6):1464–1471
12. Haller G (2001) Distinguished material surface and coherent structures in three-dimensional flows. *Physica D* 149:248–277
13. Berdahl CH, Thompson DS (1993) Education of swirling structure using the velocity gradient tensor. *AIAA J* 31(1):97–103
14. Chong MS, Perry AE, Cantwell BJ (1990) A general classification of three-dimensional flow fields. *Phys Fluids A* 2(5):765–777

15. Hunt J, Wray A, Moin P (1988) Eddies, stream, and convergence zones in turbulent flows. Technical report CTR-S88, Center for Turbulence Research, Stanford University
16. Jeong J, Hussain F (1995) On the identification of a vortex. *J Fluid Mech* 285:69–94
17. Levy Y, Degani D, Seginer A (1990) Graphical visualization of vortical flows by means of helicity. *AIAA J* 28(8):1347–1352
18. Miura H, Kida S (1997) Identification of tubular vortices in turbulence. *J Phys Soc Jpn* 66(5):1331–1334
19. Strawn RC, Kenwright DN, Ahmad J (1999) Computer visualization of vortex wake systems. *AIAA J* 37(4):511–512
20. Jiang M, Machiraju R, Thompson DS (2002) A novel approach to vortex core region detection. In: Joint eurographics IEEE TCVG symposium visualization, pp 217–225
21. Peikert R, Roth M (1999) The parallel vectors operator—a vector field visualization primitive. In: Proceedings of the 10th IEEE visualization conference (VIS 99) (Washington, DC, USA), IEEE Computer Society, pp 263–270
22. Roth M, Peikert R (1998) A higher-order method for finding vortex core lines. *IEEE Vis* 98:143–150
23. Reinders F, Sadarjoen IA, Vrolijk B, Post FH (2002) Vortex tracking and visualization in a flow past a tapered cylinder. *Comput Graph Forum* 21:675–682
24. Sujudi D, Haines R (1995) Identification of swirling flow in 3D vector fields. In: AIAA 12th computational fluid dynamics conference, paper 95-1715
25. Weinkauff T, Sahner J, Theisel H, Hege HC (2007) Cores of swirling particle motion in unsteady flows. *IEEE Trans Vis Comput Graph* 13(6):1759–1766
26. Burger R, Muigg P, Ilcik M, Doleisch H, Hauser H (2007) Integrating local feature detectors in the interactive visual analysis of flow simulation data. In: Joint eurographics IEEE TCVG symposium visualization, pp 171–178
27. Banks DC, Singer BA (1995) A predictor-corrector technique for visualizing unsteady flow. *IEEE Trans Vis Comput Graphics* 1(2):151–163
28. Stegmaier S, Rist U, Ertl T (2005) Opening the can of worms: an exploration tool for vortical flows. *IEEE Vis* 05:463–470
29. Tricoche X, Garth C, Kindlmann G, Deines E, Scheuermann G, Ruetten M, Hansen C (2004) Visualization of intricate flow structures for vortex breakdown analysis. *IEEE Vis* 04:187–194
30. Graftieaux L, Michard M, Grosjean N (2001) Combining PIV, POD and vortex identification algorithms for the study of unsteady turbulent swirling flows. *Measur Sci Technol* 12:1422–1429
31. Jespersen D, Levit C (1991) Numerical simulation of flow past a tapered cylinder. In: 29th Aerospace sciences meeting. AIAA paper 91-0751
32. Ekaterinaris J, Schiff L (1990) Vortical flows over delta wings and numerical prediction of vortex breakdown. In: AIAA aerospace sciences conference. AIAA paper, 90-0102

Applying Subgroup Discovery Based on Evolutionary Fuzzy Systems for Web Usage Mining in E-Commerce: A Case Study on OrOliveSur.com

C. J. Carmona, M. J. del Jesus and S. García

Abstract In data mining, the process of data obtained from users history databases is called Web usage mining. The main benefits lie in the improvement of the design of Web applications for the final user. This paper presents the application of subgroup discovery (SD) algorithms based on evolutionary fuzzy systems (EFSs) to the data obtained in an e-commerce Web site of extra virgin olive oil sale called <http://www.orolivesur.com>. For this purpose, a brief description of the SD process (objectives, properties, quality measures) and EFSs is presented. A discussion about the results obtained will also be included, especially focusing on the interests of the designer team of the Web site, providing some guidelines for improving several aspects such as usability and user satisfaction.

Keywords Evolutionary fuzzy system · Subgroup discovery · Web usage mining · www.OrOliveSur.com

1 Introduction

Application of data mining techniques in order to extract knowledge in a Web site automatically was considered by Etzioni [1] as Web mining which was classified in three domains with respect to the nature of data [2]: Web content mining, Web structure data, and Web usage mining.

C. J. Carmona (✉) · M. J. d. Jesus · S. García
Department of Computer Science, Building A3, University of Jaen, 23071 Jaen, Spain
e-mail: ccarmona@ujaen.es

M. J. d. Jesus
e-mail: mjjesus@ujaen.es

S. García
e-mail: sglopez@ujaen.es

This paper is focused on the extraction of useful information from Web usage data acquired using Google Analytics toolkit in the Web site <http://www.orolivesur.com>, i.e., Web usage mining applied in an e-commerce Web site. The extraction process is performed through subgroup discovery (SD) algorithms, in particular, algorithms based on evolutionary fuzzy systems (EFSs): SDIGA, MESDIF, and NMEEF-SD.

SD [3, 4] is a broadly applicable data mining technique aimed at discovering interesting relationships between different objects in a set with respect to a specific property which is of interest to the user the target variable. The patterns extracted are normally represented in the form of rules and called subgroups.

In a previous work, [5] were analyzed concepts concerning to the access properties of the users as browser or source in <http://OrOliveSur.com> in the year 2010. However, in this paper, we perform an analysis related to search engine access with respect to three values: olive oil, organic, and brand, in the year 2012 from January to June. The main objective is to obtain information related to the time and pages visited by users depending on the access keyword.

Structure of this paper is organized as follows: Sect. 2 presents the SD data mining technique: definition, properties, quality measures, and algorithms used in this study, Sect. 3 presents the main information about the e-commerce Web site in which is based on this paper “<http://www.OrOliveSur.com>,” in Sect. 4, the complete experimental study is presented, and finally, Sect. 5 presents concluding remarks about this study to the experts.

2 Subgroup Discovery

The main objective of SD task is to extract descriptive knowledge from the data concerning a property of interest [6, 7], where the main aim of these tasks is to understand the underlying phenomena with respect to an objective class and not to classify new instances.

In the following subsections, the properties of the SD task, quality measures, and algorithms used in this paper are depicted.

2.1 Properties

The concept of SD was initially introduced by Kloesgen [3] and Wrobel [4], and it can be defined as [8]:

In subgroup discovery, we assume we are given a so-called population of individuals (objects, customer, ...) and a property of those individuals we are interested in. The task of subgroup discovery is then to discover the subgroups of the population that are statistically “most interesting”, i.e., are as large as possible and have the most unusual statistical (distributional) characteristics with respect to the property of interest.

The main purpose of SD is based on the search of relations between different properties or variables with respect to a target variable, where it is not necessary to obtain complete but partial relations. Therefore, this technique uses the descriptive induction through supervised learning.

Relations are described in the form of individual rules. Then, a rule (R), which consists of an induced subgroup description, can be formally defined as [9]:

$$R : \text{Cond} \rightarrow \text{Target}_{\text{value}}$$

where $\text{Target}_{\text{value}}$ is a value for the variable of interest (target variable) for the SD task, and Cond is commonly a conjunction of features (attribute-value pairs) which is able to describe an unusual statistical distribution with respect to the $\text{Target}_{\text{value}}$.

Main elements for a SD approach are as follows: type of the target variable, description language, search engine, and quality measures. The study and configuration of these elements are very important in order to develop a new approach for SD task. In the following subsection, quality measures employed in this experimental study are shown.

2.2 Quality Measures

This element is key for extraction of knowledge because quality measures guide search process and allow to quantify quality of extracted knowledge. In addition, they show to the experts the quality of subgroups obtained. Throughout the specialized bibliography have been presented a wide number of quality measures [3, 7, 9–11]. Below, quality measures employed in this experimental study are presented:

- *Number of rules (n_r)*: This measures the number of induced rules.
- *Number of variables (n_v)*: This quality measure obtains the average of variables in the antecedent. The number of variables of the antecedent for a set of rules is computed as the average of the variables for each rule of that set.
- *Significance*: This measure indicates the significance of a finding, if measured by the likelihood ratio of a rule [3].

$$\text{Sign}(R) = 2 \cdot \sum_{k=1}^{n_c} n(\text{Target}_{\text{value}k} \cdot \text{Cond}) \cdot \log \frac{n(\text{Target}_{\text{value}k} \cdot \text{Cond})}{n(\text{Target}_{\text{value}k}) \cdot p(\text{Cond})} \quad (1)$$

where n_c is the number of values of the target variable, $n(\text{Target}_{\text{value}} \cdot \text{Cond})$ is the number of examples which satisfy the conditions and also belong to the value for the target variable, $n(\text{Target}_{\text{value}})$ is the number of examples for the target variable, and $p(\text{Cond}) = \frac{n(\text{Cond})}{n_s}$ is used as a normalized factor, n_s is the number of examples, $n(\text{Cond})$ is the number of examples which satisfy the conditions. It must be noted that although each rule is for a specific $\text{Target}_{\text{value}}$,

the significance measures the novelty in the distribution impartially, for all the values.

- *Unusualness*: This measure is defined as the weighted relative accuracy of a rule [12]. It can be computed as:

$$\text{Unus}(R) = \frac{n(\text{Cond})}{n_s} \left(\frac{n(\text{Target}_{\text{value}} \cdot \text{Cond})}{n(\text{Cond})} - \frac{n(\text{Target}_{\text{value}})}{n_s} \right) \quad (2)$$

The unusualness of a rule can be described as the balance between the coverage of the rule $p(\text{Cond}_i)$ and its accuracy gain $p(\text{Target}_{\text{value}} \cdot \text{Cond}) - p(\text{Target}_{\text{value}})$.

- *Sensitivity*: This measure is the proportion of actual matches that have been classified correctly [3]. It can be computed as:

$$\text{Sens}(R) = \frac{\text{TP}}{\text{Pos}} = \frac{n(\text{Target}_{\text{value}} \cdot \text{Cond})}{n(\text{Target}_{\text{value}})} \quad (3)$$

where *Pos* are all the examples of the target variable ($n(\text{Target}_{\text{value}})$). This quality measure was used in [13] as *Support based on the examples of the class* and used to evaluate the quality of the subgroups in the receiver operating characteristic (ROC) space. Sensitivity combines precision and generality related to the target variable.

- *Confidence*: It measures the relative frequency of examples satisfying the complete rule among those satisfying only the antecedent. This can be computed with different expressions, e.g., [14]:

$$\text{Conf}(R) = \frac{n(\text{Target}_{\text{value}} \cdot \text{Cond})}{n(\text{Cond})} \quad (4)$$

In this paper, we use fuzzy confidence [13]. It is an expression adapted for fuzzy rules which are generated by algorithms used in this experimental study.

2.3 Evolutionary Fuzzy Systems

A EFS is basically a fuzzy system augmented by a learning process based on evolutionary computation, which includes genetic algorithms, genetic programming, and evolutionary strategies, among other evolutionary algorithms [15]. Fuzzy systems are one of the most important areas for the application of the fuzzy set theory [16]. Usually, this kind of systems considers a model structure in the form of fuzzy rules. They are called fuzzy rule-based systems (FRBSs), which have demonstrated their ability with respect to different problems like control problems, modeling, classification, or data mining in a large number of applications. FRBSs provide us a comprehensible representation of the extracted knowledge and moreover a suitable tool for processing the continuous variables.

Three algorithms based on EFSs for SD task have been presented:

- SDIGA is an evolutionary model for the extraction of fuzzy rules for SD [13]. The use of a knowledge representation based on fuzzy logic and the use of evolutionary computation as a learning process receive the name of evolutionary fuzzy system [17]. This type of systems for SD task has been applied in different real-world applications like [18–20].
- MESDIF is a multi-objective evolutionary algorithm for the extraction of fuzzy rules which describe subgroups [21]. The algorithm extracts a variable number of different rules expressing information on a single value of the target variable. The multi-objective evolutionary algorithm is based on the SPEA2 approach and so applies the concepts of elitism in the rule selection (using a secondary or elite population) and the search for optimal solutions in the Pareto front. In order to preserve the diversity at a phenotypic level, the algorithm uses a niches technique which considers the proximity in values of the objectives and an additional objective based on novelty to promote rules which give information on examples not described by other rules of the population.
- NMEEF-SD [22] provides from Non-dominated Multi-objective Evolutionary algorithm for extracting fuzzy rules in SD. This is an evolutionary fuzzy system whose objective is to extract descriptive fuzzy and/or crisp rules for the SD task, depending on the type of variables present in the problem. NMEEF-SD has a multi-objective approach based on NSGA-II, which is a computationally fast MOEA based on a non-dominated sorting approach, and on the use of elitism. The proposed algorithm is oriented toward SD and uses specific operators to promote the extraction of simple, interpretable, and high quality SD rules. The proposal permits a number of quality measures to be used both for the selection and for the evaluation of rules within the evolutionary process.

3 OrOliveSur.com an E-commerce Website Related to Organic Extra Virgin Olive Oil

OrOliveSur is a project born in the province of Jaén from Andalusia (Spain) in 2010. The main purpose is to announce to the world the treasure of its land, the extra virgin olive oil. This Web site is focused in the olive oil produced in a particular territory of Jaén: the Sierra Mágina Natural Park. Sierra Mágina is a protected area of 50,000 acres of natural park, made up of forested slopes, concealed valleys, and rugged mountain peaks. The highest peak, the Mágina Mountain is the highest in the Jaén province, standing at 2,167 m.

OrOliveSur's catalog presents a wide number of extra virgin olive oils focused on the Picual variety. This is the most extended olive grove variety at the world. In Spain, it represents 50 % of production. Most of it is to be found in Andalusia, especially in the province of Jaén. Its olive is large sized and elongated in shape,

with a peak at the end. The trees of this variety are of an intense silvery color, open, and structured.

Along 2 years, OrOliveSur has received both national and international orders from European Union countries (Spain, Denmark, Germany, Great Britain, France, etc.), and its visits and orders are increased every day. Moreover, the OrOliveSur

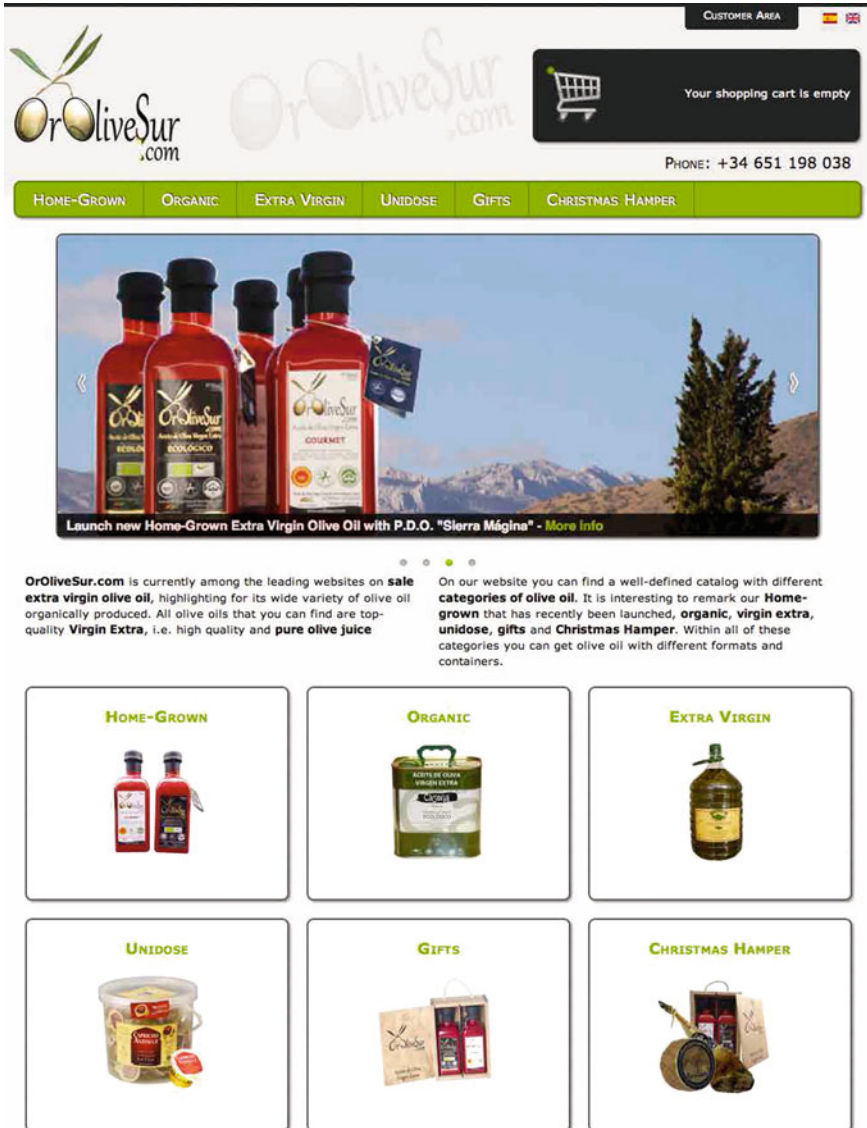


Fig. 1 Homepage from the e-commerce Web site <http://OrOliveSur.com>

Web site gives direct sales and clients can pay by transfer bank, PayPal, or credit card. In Fig. 1, the homepage of OrOliveSur is shown.

4 Experimental Study: Web Usage Mining Applied to OrOliveSur.com

The experimental study presented in this paper is focused on Web usage mining which was defined by Srivastava [23] as:

The process of applying data mining techniques to the discovery of usage patterns from Web data.

Patterns are represented as a collection of pages or items visited by users. These patterns can be employed to understand the main features of the visitants behaviors in order to improve the structure of one Web site and create personal or dynamic recommendations about content of the web. The main purpose is to analyze the interaction of the users in the e-commerce Web site of <http://OrOliveSur.com> through Web usage mining techniques. With results obtained, recommendations about the design or the access of the users could be given in order to improve the e-commerce Web site.

Next, experimental framework is presented in Sect. 4.1, and results obtained are shown in Sect. 4.2.

4.1 Experimental Framework

Database has been obtained with the Webmaster tool *Google Analytics* from the period January 1 to June 30, 2012. Moreover, several filters have been applied in data set in order to obtain only instances where bounce rate is lower than 100.00 %. This value is the percentage of single-page visits or visits in which the person left your site from the landing page, i.e., we only obtain visits where users have been visited the Web site more than one seconds. In total, the data set is composed by 2.340 instances.

Variables analyzed in this experimental study are described below:

- **Keyword:** It is the keyword access to the Web site by the user. The complete keyword set has been classified in three categories. It is important to remark that keywords of the original data set can be found in different languages, and they are classified in a general category with English terms: olive oil, organic, brand. This variable is used as target variable in the experimental study.
 - Olive oil. This value contains all the generic keywords related to the olive oil like *buy olive oil*, *venta de aceite*, *organic olive oil*, *aceite de oliva virgen extra*, *huile d'olive*, *Oliven öl Extra Vergine*, and so on.

- Brand. This keyword contains the entries related to any brand of the catalog as *La Casona*, *Verde Salud*, *Gamez-Piñar* or *OrOliveSur*, for example.
- Organic. This keyword contains all generic keywords related to organic olive oil like *organic olive oil*, *buy organic olive oil*, and so on.
- New visitor (NV): It indicates if the user is a new or returning visitor.
- Page views (PV): It indicates the page views for users with the same browser, visitor type, keyword, and source.
- Unique page views (UPV): It presents the number of unique page views by users with the same browser, visitor type, keyword, and source.
- Time on site (TS): This feature indicates the time spent on Web site by users with the same browser, visitor type, keyword, and source.
- Average time on page (ATP): It shows the average time used by the user per page view.

In Table 1 are described parameters used by EFSs for SD described previously.

4.2 Results Obtained

In this section are presented results obtained with respect to the data set <http://OrOliveSur.com>. Table 2 shows the most important rules obtained in this experimental study. In addition, values of quality measures are presented in order to understand the quality of these rules.

In the experimental study have been obtained 40 rules in total for all algorithms. However, in this section have presented the most interesting rules for each algorithm executed. In summary, two rules for SDIGA algorithm (one for keyword \rightarrow brand and keyword \rightarrow olive oil), three for MESDIF algorithm (one for each value of the target variable) and two rules for NMEEF-SD (one for keyword \rightarrow brand and keyword \rightarrow olive oil).

Table 1 Parameters of algorithms employed

Algorithm	Parameters
SDIGA	Population size = 100, evaluations = 10,000, crossover probability = 0.60, mutation probability = 0.01, minimum confidence = 0.4, representation of the rule = canonical, linguistic labels = 3, objective1 = sensitivity, objective2 = unusualness, objective3 = fuzzy confidence, weight for objective1 = 0.4, weight for objective2 = 0.3 and weight for objective3 = 0.3
MESDIF	Population size = 100, evaluations = 10,000, elite population = 3 individuals, crossover probability = 0.60, mutation probability = 0.01, representation of the rule = canonical, linguistic labels = 3, objective1 = sensitivity and objective2 = unusualness
NMEEF-SD	Population size = 50, evaluations = 10,000, crossover probability = 0.60, mutation probability = 0.1, minimum confidence = 0.4, representation of the rule = canonical, linguistic labels = 9, objective1 = sensitivity and objective2 = unusualness

Table 2 Rules and results obtained by EFSs in this case of study

Algorithm	#	Rule	Significance	Unusualness	Sensitivity	Fuzzy confidence
SDIGA	R1	IF TS = Medium THEN keyword = brand	8.285	0.007	0.997	0.466
	R2	IF TS = Low THEN keyword = olive oil	0.604	0.006	0.995	0.450
MESDIF	R3	IF TS = Medium THEN keyword = brand	8.285	0.007	0.997	0.466
	R4	IF PV = Low THEN keyword = olive oil	0.001	0.003	0.444	0.444
	R5	IF UPV = Low AND ATP = Medium THEN keyword = organic	0.233	0.001	0.232	0.232
NMEEF-SD	R6	IF TS = Medium THEN keyword = brand	8.285	0.007	0.997	0.466
	R7	IF PV = Low THEN keyword = olive oil	0.001	0.003	0.444	0.444

As can be observed in Table 2, there is a coincidence between the algorithms because rules *R1*, *R3*, and *R6* are the same rule. This rule represents users that remain in the Web site during an interesting period of time in the Web site. These users accessed to OrOliveSur through a keyword related to a brand. However, with respect to the keyword olive oil, this variable obtains the linguistic label *Low*, i.e., when users access to the Web site through a keyword introduced in a search engine related to olive oil, they land in the Web site but they remain in small period. With rules *R2* and *R7* obtained by SDIGA and NMEEF-SD, respectively, we could consider that users cannot find information that they expected. Results obtained for the keyword organic are not very precise because in the database, there are few instances for this value. However, as can be observed in rule *R5* is interesting to remark that despite of the number of page views is low, users remain in the Web site during a good period because the average time on page is medium.

With respect to the results obtained for each fuzzy subgroup, on the one hand, a good sensitivity for subgroups obtained for keyword brand with values close to 100 % and high values of significance can be observed. On the other hand, fuzzy confidence values are close to 50 % in keywords olive oil and brand. It is interesting to remark the use of few variables in subgroups to describe the problem.

5 Conclusions

This study presents the application of SD algorithm-based EFSs to the real-world application OrOliveSur: an e-commerce Web site related to olive oil and organic olive oil. The main objective is to extract unusual knowledge about users history associated with the Web site. SD algorithms allow extract knowledge with respect

to a target variable, where it is not necessary to obtain complete but partial relations. In this way, SD uses the descriptive induction through supervised learning.

The most interesting fuzzy subgroups obtained by SDIGA, MESDIF, and NMEEF-SD are presented. Conclusions proportioned to the Webmaster team with respect to the knowledge extracted are:

- The design must be reviewed with respect to the appearance and text of products in order to follow than users remain in the Web site during more time.
- A new categorization should be generated in order to facilitate the users different types of olive oils included in the Web site. In this way, users could explore more pages and perform more orders.
- Search Engine Optimization with respect to the keyword organic must be performed because there are few visits with respect to brand.

Acknowledgments This paper was supported by the Spanish Ministry of Education, Social Policy and Sports under project TIN-2008-06681-C06-02, FEDER Funds, by the Andalusian Research Plan under project TIC-3928, FEDER Funds, and by the University of Jaén Research Plan under project UJA2010/13/07 and Caja Rural sponsorship.

References

1. Etzioni O (1996) The World Wide Web: quagmine or gold mine. *Commun ACM* 39:65–68
2. Cooley R, Mobasher B, Srivastava J (1997) Web mining: information and pattern discovery on the World Wide Web. On tools with, artificial intelligence, pp 558–567
3. Kloesgen W (1996) Explora: a multipattern and multistrategy discovery assistant. In: *Advances in knowledge discovery and data mining*. American Association for, artificial intelligence, pp 249–271
4. Wrobel S (1997) An algorithm for multi-relational discovery of subgroups. In: *Proceedings of the 1st European symposium on principles of data mining and knowledge discovery*. Volume 1263 of LNAI. Springer, New York, pp 78–87
5. Carmona CJ, Ramírez-Gallego S, Torres F, Bernal E, del Jesus MJ, García S (2012) Web usage mining to improve the design of an e-commerce website: OrOliveSur.com. *Expert Syst Appl* 39:11243–11249
6. Lavrac N, Kavsek B, Flach PA, Todorovski L (2004) Subgroup discovery with CN2-SD. *J Mach Learn Res* 5:153–188
7. Herrera F, Carmona CJ, González P, del Jesus MJ (2011) An overview on subgroup discovery: foundations and applications. *Knowl Inf Syst* 29(3):495–525
8. Wrobel S (2001) Relational data mining. In: *Inductive logic programming for knowledge discovery in databases*. Springer, New York, pp 74–101
9. Lavrac N, Cestnik B, Gamberger D, Flach PA (2004) Decision support through subgroup discovery: three case studies and the lessons learned. *Mach Learn* 57(1–2):115–143
10. Gamberger D, Lavrac N (2003) Active subgroup mining: a case study in coronary heart disease risk group detection. *Artif Intell Med* 28(1):27–57
11. Kloesgen W, Zytkow J (2002) *Handbook of data mining and knowledge discovery*. Oxford
12. Lavrac N, Flach PA, Zupan B (1999) Rule evaluation measures: a unifying view. In: *Proceedings of the 9th international workshop on inductive logic programming*. Volume 1634 of LNCS. Springer, New York, pp 174–185

13. del Jesus MJ, González P, Herrera F, Mesonero M (2007) Evolutionary fuzzy rule induction process for subgroup discovery: a case study in marketing. *IEEE Trans Fuzzy Syst* 15(4):578–592
14. Agrawal R, Mannila H, Srikant R, Toivonen H, Verkamo A (1996) Fast discovery of association rules. In Fayyad U, Piatetsky-Shapiro G, Smyth P, Uthurusamy R (eds.) *Advances in knowledge discovery and data mining*. AAAI Press, pp 307–328
15. Eiben AE, Smith JE (2003) *Introduction to evolutionary computation*. Springer, New York
16. Zadeh LA (1975) The concept of a linguistic variable and its applications to approximate reasoning. Parts I, II, III. *Inf Sci* 8–9:199–249,301–357,43–80
17. Herrera F (2008) Genetic fuzzy systems: taxonomy, current research trends and prospects. *Evol Intell* 1:27–46
18. Carmona CJ, González P, del Jesus MJ, Romero C, Ventura S (2010) Evolutionary algorithms for subgroup discovery applied to e-learning data. In: *Proceedings of the IEEE international education, engineering*, pp 983–990
19. Carmona CJ, González P, del Jesus MJ, Navío M, Jiménez L (2011) Evolutionary fuzzy rule extraction for subgroup discovery in a Psychiatric Emergency Department. *Soft Comput* 15(12):2435–2448
20. Carmona CJ, González P, del Jesus MJ, Ventura S (2011) Subgroup discovery in an e-learning usage study based on Moodle. In: *Proceedings of the international conference of European transnational, education*, pp 446–451
21. del Jesus MJ, González P, Herrera F (2007) Multiobjective genetic algorithm for extracting subgroup discovery fuzzy rules. In: *Proceedings of the IEEE symposium on computational intelligence in multicriteria decision making*. IEEE Press, pp 50–57
22. Carmona CJ, González P, del Jesus MJ, Herrera F (2010) NMEEF-SD: non-dominated multi-objective evolutionary algorithm for extracting fuzzy rules in subgroup discovery. *IEEE Trans Fuzzy Syst* 18(5):958–970
23. Srivastava J, Cooley R, Deshpande M, Tan P (2000) Web usage mining: discovery and applications of usage patterns from web data. *SIGKDD Explorations*, pp 12–23

Distributed Data Distribution Mechanism in Social Network Based on Fuzzy Clustering

Yan Cao, Jian Cao and Minglu Li

Abstract With the development of Internet, especially the success of social media like Facebook, Renren and Douban, social network has become an important part of people's life. For social applications, one of the key problems to be solved is how to distribute data accurately to different users and groups in high speed. In this paper, we introduce fuzzy clustering into social network analysis. Users with similar interests are clustered into the same network according to fuzzy similarities. Our goal is to study how such clustering can enhance data distribution to the largest extent. We take 3,000 user's data from Douban.com, fetch fuzzy social relationship, and simulate data transmission with a theme-based pub/sub mechanism. Experiments show that network clustering based on fuzzy clustering can improve data distribution effectively, while remain robust in highly dynamic environment.

Keywords Social network · Data distribution · Fuzzy clustering · Network clustering · Social computing

1 Introduction

With the development of Internet and people's ever increasing demands for information sharing, there comes lots of data distribution applications. For these applications, how to distribute data accurately to users with different requirements

Y. Cao (✉) · J. Cao · M. Li

Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China

e-mail: yancao1122@gmail.com

J. Cao

e-mail: cao-jian@cs.sjtu.edu.cn

M. Li

e-mail: li-ml@cs.sjtu.edu.cn

is a key problem. Especially for social network service, its highly dynamic environment makes it harder to transmit data in a fast and adaptive manner. Currently, most data distribution service in social network can be divided into two kinds: concentrated service and distributed service.

In concentrated network service, all user information and social relationship are stored and maintained with concentrated servers. The data distribution process is also realized with concentrated servers. Most such services support both real-time data transmission and off-line data reception. Mainstream social network services now, for example, Facebook, Twitter, Myspace, and Douban, are built with concentrated services. However, problems related to privacy, extensibility, and single-point inactive are naturally with concentrated services. Therefore, distributed social network service becomes a focal point. It is usually a P2P service without concentrated management and control. User data are usually stored locally in reliable neighbor node. Therefore, privacy, extensibility, and robustness are guaranteed to some extent. However, due to its own characteristics and highly dynamic user behavior in social network, most current distributed data distribution applications for social network do not realize functions provided by concentrated services. While at the same time, pub/sub system, born for dynamic data and various data distribution requirements, became an ideal framework because of its loose coupling feature. In comparison with traditional communication models (message transmission, RPC/RMI, and shared space), pub/sub enjoys the advantages of asynchronous and multi-point communication, enabling participants to decouple entirely regarding space, time, and control flow, so as to facilitate loose communication in large-scale distributed system. For social network, data distribution is consisted of mostly messages between friends and groups. Especially for each user, a group equals a theme in theme-based pub/sub system.

However, pub/sub itself is only a framework. How to satisfy dynamic and robust data distribution demands is still a critical issue applying it. A typical pub/sub model can be divided into three parts: topology, events routing, and events matching. Existing content-based data distribution mechanism takes on different kinds of topology and routing model. These models, however, all use random routing, which cannot guarantee transfer efficiency in many cases. Obviously, in social network with fruitful social resources, constructing topology with analyzed social information is more efficient than simple random network. This paper is to address such a problem: how to use social information to cluster network, that is, organize users into different clusters, so as to efficiently improve data distribution efficiency to the largest extent.

When we study social relationship, the first problem to resolve is how to represent it in a concise and significant way. One common method is to use two-dimensional adjacency matrix to represent connections. It is weak, however, since it can only show relationship in pair without strength of this relationship. Nor can it disclose the relationship's social or personal attributes. Introducing fuzzy social relationship into SNS, however, naturally overcomes this problem perfectly and extends social definition. Fuzzy graph is used to represent social network, taking into account both continuous changes in dynamic social environment and

user's quantitative attributes. So far, there has not much research in this respect, even no combining network clustering with fuzzy relationship definition. This paper's uniqueness lies here: represent existing social information, for example, user's interests, activities, and friends with fuzzy definition, calculate users' fuzzy similarity with fuzzy clustering analysis, and classify users with similar interests into the same network cluster according to ontology library with fuzzy clustering analysis. The merging and combination of route table are based on ontology library, which evolves together with dynamically changed network. For each user, we define a membership function to represents its connection with certain cluster, so that network clustering, user distribution, and data transfer can adapt to dynamic changes in user behavior, ensuring consistent high data distribution efficiency.

To evaluate impact of fuzzy clustering-based network on data distribution, this paper uses a theme-based pub/sub system to help. Because people naturally transfer events with their interested themes, such clustering enables events with similar themes to be transferred always within the same network. Since most data transfer is done within network, instead of across, average hop is reduced a lot, and data distribution efficiency is enhanced significantly. Transfer time is shortened, and bandwidth resources are also saved. This paper's major contribution is as follows:

As the first work to introduce fuzzy clustering into social network analysis, we establish ontology library and cluster network according to it, so as to organize users who like the same theme into a same network. We classify the relationships of different social network links based on a small subset of known social relationships data.

By taking different relationships into account, we further investigate the impact of fuzzy clustering network on data transmission. We employ theme-based pub/sub model to perform data routing tasks.

We extensively evaluate our model on a large-scale online social network. To this end, we have collected more than 3,000 users' data from Douban network for our evaluation. Experiment shows that our fuzzy clustering methodology fits into dynamic social network. In comparison with random network and hash clustering network, our clustering model brings higher transfer efficiency and lower delay.

The next sections are organized as follows. In section II, we present some preliminaries of our research. In section III, we elaborate on details of our fuzzy clustering model of social network, including its fuzzy representation and how it may evolve with changes in the network. In section IV, we clarify the theme-based pub/sub matching and routing mechanism, which facilitates us in data transmission and distribution. We then conduct an experiment to evaluate our system's performance and present the results in section V, comparing this system's event receive rate and average hop times with random distributed network and hash-table-based multi-level networks. We conclude our work in section VI and related work and references in the last section.

2 Preliminaries

In this section, we will present the social network models used and formally define the problems to be studied in this paper.

2.1 Social Network Model

An online social network is often composed of users, links, and groups. As in all online social networks, to participate fully in an online social network, a user (often a human being) must register with the site. After a user registered in a site, the user can create links to other users in the same social network. Here, users form links for various reasons: the users can be real-world acquaintances or business contacts; they can share some common interests; or they are interested in each other's contents. For a user u , the set of users with whom u has links are called the *contacts* of the user u . Most sites enable users to create and join special interest groups. Users can post messages to groups and upload shared content to the group. In many of these sites, links between users are public and can be crawled automatically to capture and study a large fraction of the connected user graph.

In this paper, we formulate a social network as a graph $G = (V, E)$, in which V is the set of users in the social network, and E is the set of links among users. We assume that the links in the social network can be classified into a set of categories $L = \{L_1, L_2, L_3, \dots, L_k\}$, such as {friends, classmates, co-workers, family, others}. In other words, each link $e \in E$ has one set of multiple labels $l(e) \in L$. Note that each label has a $K(l)$ to notify its degree of belonging to the link.

Since the social network is highly dynamically changing, we just incorporate its dynamic nature into our notification. Let $G^t = (V^t, E^t)$ be a time-dependent network snapshot recorded at time t . Let ΔV^t and ΔE^t represent the sets of vertices and links to be introduced (or removed) at time t , and let $\Delta G^t = (\Delta V^t, \Delta E^t)$ denote the change in terms of the whole network. The next network snapshot G^{t+1} is the current one together with changes, that is, $G^{t+1} = G^t \cup \Delta G^t$. A dynamic network G is a sequence of network snapshots evolving over time: $G = (G^0, G^1, G^2, \dots, G^t)$.

2.2 Problem Definition

An explicit assumption made in many existing works on social network mining is that the social network G typically contains only one relationship; that is, the relationships among different users are identical as long as they have a link between each other. However, as in a real social network, a user can simultaneously have relationships with several topics and each with different degrees of belonging, and we donate such relationship in a fuzzy manner and establish a one-to-many relationship. Given a social network $G = (G^0, G^1, G^2, \dots, G^t)$ where G^0

is the original network, and G^1, G^2, \dots, G^t are the network snapshots obtained through $\Delta G^1, \Delta G^2, \dots, \Delta G^t$, we need to devise an adaptive algorithm to efficiently classify relationship in the network and cluster the network into a multi-level tree structure at any time point utilizing the information from the previous snapshots. We also employ this clustered structure in a social network and explore how it may influence data transmission performance of this network.

3 Fuzzy Clustering of Social Information

In this section, we will study how to classify the relationship in online social networks efficiently and effectively. We will mainly focus on learning the social relationships of pairs of nodes in the social network starting from fuzzy representation of social relationships.

3.1 Fuzzy Relationship Representation

As already mentioned, SNA is the branch of network analysis devoted to studying and representing relationships between “social” objects. To formalize, SNA mainly explores relationships between objects belonging to an universal set $X = \{x_1, \dots, x_n\}$, and in order to achieve its aim, some mathematical properties of relationship are utilized. More specifically, a binary relation on a single set, which is the most popular kind of relation used in the SNA, is a relationship $A \subseteq X \times X$, whose characteristic function

$$\mu_A(x_i, x_j) = \begin{cases} 1, & \text{if } x_i \text{ is related to } x_j \\ 0, & \text{if } x_i \text{ is not related to } x_j \end{cases} \quad (1)$$

Some scholars in the field claim that A has its strong point in being a good synthesis of all the pairwise relationship between elements of X . In contrast, according to some others, A is too poor of information, that is, it does not contain information about the degree to which the relationships between two elements hold. Therefore, it may happen that it treats in the same way very different cases, without discriminating among situations where intensities of relationship may be very different. Indeed, many examples may be brought in order to support the latter point of view.

Some methods have already been proposed in order to overcome the problem related to the lack of information about the intensity of relationship between elements of a pair. For instance, a discrete scale can be adopted and a value be assigned to each entry a_{ij} to denote the intensity of relation between x_i and x_j . This approach, based on *valued adjacency relations*, is the most widely used in order to overcome the problem of unvalued relations.

Here, we want to propose an alternate approach based on *fuzzy sets* theory in order to obtain a *fuzzy* adjacency relation. A binary fuzzy relationship on a single set, $R_2 \subseteq X \times X$, is defined through the following membership function

$$\mu_{R_2} : X \times X \rightarrow [0, 1], \quad (2)$$

and also in this case, putting $rij := \mu_{R_2}(xi, xj)$, a fuzzy relationship can be conveniently represented by a matrix $R = (rij)_{n \times n}$, where the value of each entry is the degree to which the relationship between xi and xj holds. In other words, the value of $\mu_{R_2} = (xi, xj)$ is the answer to the question: "How strong is the relationship between xi and xj ?" Therefore, in the context of SNA

$$\mu_{R_2} = \begin{cases} 1, & \text{if } xi \text{ has the strongest possible degree of relationship with } xj \\ \gamma = [0, 1], & \text{if } xi \text{ is, to some extent, related to } xj \\ 0, & \text{if } xi \text{ is, not related with } xj \end{cases} \quad (3)$$

Fuzzy adjacency relationships, as well as crisp adjacency relations, are here assumed to be reflexive and symmetric. We can shift from the fuzzy approach to the crisp one thanks to the ∞ -cuts. An ∞ -cut is a crisp relation defined by

$$\mu_{R_2}(xi, xj) = \begin{cases} 1, & \text{if } \mu_{R_2}(xi, xj) \geq \infty \\ 0, & \text{if } \mu_{R_2}(xi, xj) < \infty \end{cases} \quad (4)$$

3.2 Common-Shared Ontology Library

We establish ontology library on high level to be an independent module. This library stores concepts and their hierarchical order in a structured way, which is useful in both network clustering and matching of information during data transmission process. It also evolves as the data transmission begins. What's more, considering the participants in the systems come from various organizations, and each ontology in the library is donated to a domain, that is, each ontology has a specific label about its domain. This helps to differentiate ontology and facilitate data routing.

In this ontology library, the ontology we defined can be expressed in a 5-tuple set: (C, P, R, F, A) . C means set of conceptions, and P is the set of conception's attribute. R shows the relationship of the ontology support. Here, we mainly define three relationships:

$$\text{DOM}(R) = \{E : \text{equivalent}, S : \text{subClassOf}, A : \text{attributeOf}\} \quad (5)$$

- *E*: equivalent ($C1, C2$) means $C1$ and $C2$ are of the same conceptual level. For example, zebra and elephant are equivalent in that they are both the top-level concepts after “animal”.
- *S*: subClassOf ($C1, C2$) means $C2$ is the subclass of $C1$. So when the subscriber is interested in $C1$, he is also interested in the child of $C1$, namely $C2$. For example, the person who is interested in “computer”, he should be interested in “notebook” and “desktop”.
- *A*: attributeOf ($C1, P$) means P is the attribute of $C1$.

F means a set of function according to the relationship. *A* is a group of predicate logic expression to express the constrain rule in ontology. Figure 1 shows the ontology structure we just mentioned.

We then move on to the data model of ontology. Multi-dimensional index structure has been widely used in many application fields such as database and image search in order to allow users to acquire information swiftly. *X* tree is a kind of multi-dimensional index structure, which combines class hierarchical structure with supernode (i.e., a size-flexible class linear array). It is suitable to express ontology relationship hierarchy in each conception. Figure 2 shows the *X* tree-based ontology. The construction process works as follows: First, each concept or property is matched to relate the domain according to its semantic meaning, and we can get the supernode through the domain link if it is a concept. The supernode is an array expressed as (*core concept*, *attribute*, *child*, *parent*), in which the *core concept* represents all the synonym in the domain, *attribute* saves

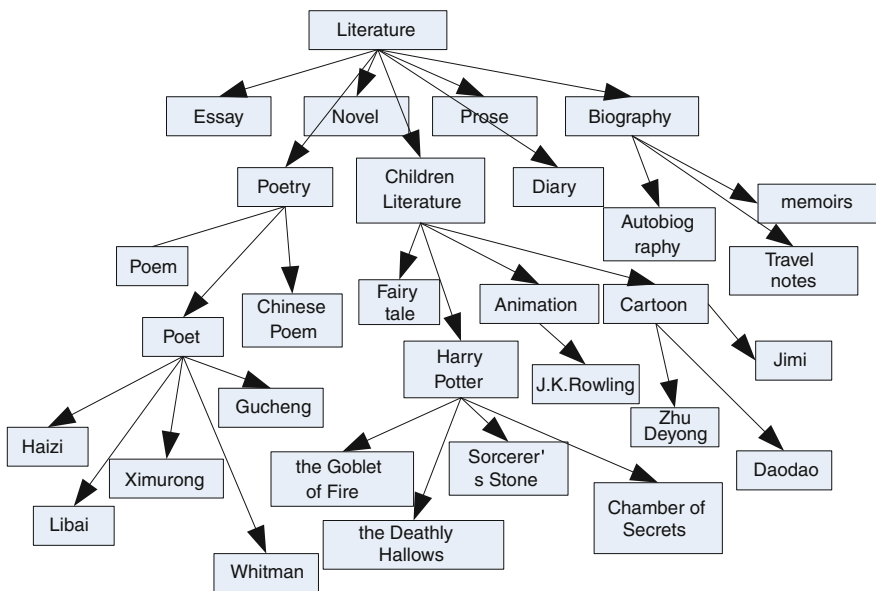


Fig. 1 Ontology library structure

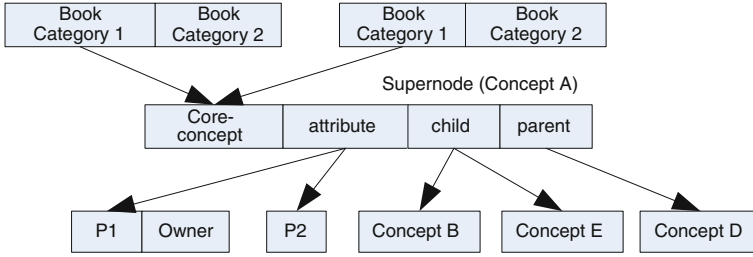


Fig. 2 Ontology node structure

all the relevant attributes it owns, *child* links to the subclass it has, and superclass links are in *parent*. The property of the concept is an array expressed as *property*, *owner*, and *belongings*: *owner* links to the owners of the property and *belongings* donate how close they connect to each other.

3.3 Network Definition

To better perform data distribution, this network is based on a multi-level graph topology, in which data points, namely users, are organized into different network clusters and clusters are maintained dynamically and adaptively with fuzzy rules. General social information, which includes various themes such as books, music, activities, or clubs are collected and classified into different categories. Since each item, for example, a book or a song, may belong to several categories simultaneously, for example, a thrilling love story or a piano masterpiece originated in Italy, it is inadequate if we attribute it into any single category. Also if we simply include it into all those related categories, there will be great amount of redundancy. To carefully solve this problem, here we propose to represent such item data with fuzzy descriptions. Take books as an example, n books' information is used to characterize the network, and m represents the number of different themes,

$$X = \{x_1, x_2, \dots, x_n\} \tag{6}$$

$$x_i = (x_{i1}, x_{i2}, \dots, x_{im}) \tag{7}$$

A fuzzy analogous matrix is then calculated to establish analogous relationships among each item.

$$r_{ij} = 1 - c \sum_{k=1}^m |x_{ik} - x_{jk}|, \quad 0 \leq c \leq 1 \tag{8}$$

We then use transitive closure clustering to classify the items. For each $R = (r_{ij})_{m \times n}$, R^2, R^4, \dots, R^{2^l} are calculated with fuzzy multiplication \circ , where for matrix A and B ,

$$C = A \circ B = (c_{ik}) = \bigvee_{j=1}^m (a_{ij} \wedge b_{jk}) \tag{9}$$

where \bigvee represents logically add, and \bigwedge represents logically minus, and when $R^k \circ R^k = R^k$ appears for the first time, $R^k = t_{ij}$ is named as a transitive closure.

For $t_{ij} = t(R), 0 \leq t_{ij} \leq 1$, let λ be one value among t_{ij} , we then get the λ -cutting matrix of $t(R)$, the column vector of $t(R)$ is related to elements in $X = \{x_1, x_2, \dots, x_n\}$, and elements in X belongs to the same cluster if and only if their related column vectors in R_λ are the same.

$$R_\lambda = \lambda_{ij}, \tag{10}$$

$$\lambda_{ij} = \begin{cases} 1, & \text{when } t_{ij} \geq \lambda \\ 0, & \text{when } t_{ij} < \lambda \end{cases} (i, j = 1, 2, \dots, n) \tag{11}$$

By carefully selecting λ , we can then cluster each book items into different clusters. However, as descriptions of books vary, the size of each cluster can be either too large or too small. We need to define a maximum size of each cluster, namely M , and also number of levels, namely L , to control the network scale. As the network dynamically changes and users subscribe or publish more themes, clusters that are larger than M will have to be clustered again, adding L to 1. In this way, we are actually building a network tree, in which each network could be a father or a sibling of neighboring clusters.

3.4 User Classification

Through carefully analyzing user information, we are trying to learn user's interests and collect those in the same interest into one network. In this way, we are getting as many possible information to be transferred into one cluster, so as to defer transportation delay and decrease energy consumption. For each user, we define a membership function as to decide to what extent he belongs to any cluster. Applying fuzzy relationships to SNA, the normalized complement index of each user is defined as follows:

$$C_A(x_i) = v_A \lim_{n \rightarrow \infty} \frac{\sum_i^n x_i \in A}{n} \tag{12}$$

where A refers to different categories that are important to be considered, that is, user's interest in book, music, activities, or celebrities, and v_A refers to the degree of importance of these categories.

In social network, where user information is maintained at the very beginning, we can have a initial calculation of $C_A(x_i)$ from the start. However, as the network

evolves, users may have different interests, and their relationship with each category could also vary. In this case, such definition of user–network relationship should be dynamic and adaptive. Therefore, we adjust this value each time when user subscribes a new theme or publishes a new event. Such a positive feedback system works like this:

$$x_i + = \begin{cases} 1, & \text{if a new subscription is related to } A \\ 1, & \text{if } x_i \text{ receive a new event related to } A \end{cases} \quad (13)$$

If a user starts a subscription about a theme that he has never subscribed before, however, the membership index of this user has to be calculated again, so as to allocate certain interest to that new network. In order to do so, every time a user subscribes a theme, we need to match it with this user’s networks so as to decide whether this new theme belongs to A .

4 Theme-Based Pub/Sub Model

In order to transmit data among levels of networks and also within certain clusters, we employ theme-based pub/sub system for data distribution. In such a scheme, end users, who are also called subscribers, subscribe information with certain theme and wait for events that contain that theme. Subscribers do not care who will send them events. Specifically in our system, it is users in the same network who are most likely to send interested events to each other. For publishers, it may come up with an event in certain theme and then transmit it through networks. Again publishers also do not care who will receive their events. In this way, such a data distribution model realizes low coupling and guarantees the independence of both subscription and publishing processes.

4.1 Participants and Roles

Despite end users, we here in this system include also inner brokers and boarder brokers to perform pub/sub matching and data distribution tasks.

Inner brokers are the mobile ego servers who transmit data both among and within networks. As our system is distributed over different locations, inner brokers of one cluster may also scatter in discreet geographic plots. In general, we need to guarantee that inner brokers of different clusters are distributed relatively balanced so as to avoid over delay in any one cluster. For example, if m inner brokers are distributed in n locations within r clusters, it is better to put u inner brokers in one location, given the density of clusters for this location is k ,

$$U = k * m / n * r \quad (14)$$

However, such balance is not fixed given the highly dynamic nature of social network. As users' interest change with time, density of clusters also changes. Therefore, we need to monitor this number carefully and let a viable v to represent how much such change accounts for. If v is larger than certain number, say 50 %, which means the density k may increase for 2–3, we need to calculate the U again and get another allocation of inner brokers. More brokers will be added to this location accordingly.

Boarder brokers are the mobile ego servers who at one end connect to clients and at the other end connect to inner brokers. The major difference between inner brokers and boarder brokers is the route table: items in inner broker route tables are also brokers, while those in boarder brokers are clients. The number and location of boarder brokers are closely related with clients. According to its computing capability, certain number of clients will be allocated to one boarder broker, which then connects to inner brokers in this location. Different from inner brokers, however, boarder brokers are often fixed to certain location range if no great mobility of users takes place.

Clients are users located within the clusters. Each client is represented by a user name in real social network. Clients can submit subscriptions, publish events, and receive subscriptions and events without any knowledge about their sources. Clients who have the same interests are often organized into the same cluster, and as clients' interests change, their degrees of belongings to each cluster also vary. Our system keeps track of clients' interests and adjusts cluster distribution when such changes accumulate to certain degree.

4.2 Theme-Based Routing Strategy

As in other pub/sub models, in this paper, publishers and subscribers sustain a loose relationship, by which data are transmitted without clear identification of its original source. What is different from traditional pub/sub is that every subscription and event is marked with certain themes. As they each may belong to several themes simultaneously with different degrees of belonging, it is actually a set of $M = \{theme, belonging\}$ that depicts the relationship. The subscription process works as follows:

- One subscription issued by a client arrives at connecting boarder broker at first. The broker carefully searches the theme in the ontology library and sees whether it matches any existing concepts. If it does, then this theme will be forwarded to that cluster and its *subclass* clusters. If it does not, however, this theme will be added to this broker's own library, and this theme is then transmitted to all clusters that are *equivalent* to this cluster.
- Since each subscription $S = \{filter_1, filter_2, \dots, filter_k\}$ and each filter $F = \{attribute, type, operator, value\}$, brokers that receive subscriptions then store them and record their previous hops in the route tables.

The publishing process works as follows:

- Once a user issues an event (i.e., a XML document), it arrives first at connecting boarder broker. As each event is composed of defined sets: $E = \{\text{attribute, type, value}\}$, the broker then matches this event with all its stored subscriptions, especially their filters. If matches are found, the broker then looks into its route table for corresponding next hop and forwards this event to it. If no matches are found, however, the event will not be forwarded.

As each user may belong to various clusters for the same time, subscriptions and events are sent to several boarder brokers. We therefore need to ensure that documents are not transmitted to the same recipients repeatedly. When one broker or client receives certain subscription or event, it stores this document's index number, which is generated uniquely. It can compare this number with upcoming other documents, and if any future document's index is the same with stored number, this document is rejected. In this way we avoid cycling in the network; however, there will be naturally some redundancy since we are not notified when any client receives subscription so as to stop transmission. However, such redundancy is not all about disadvantages; it to some extent avoids lost of documents due to sudden changes in the network, that is, sudden halt of power or invalid players. Moreover, as subscription continues to arrive in one broker, it automatically merges items in the route tables to make it more compact and efficient. Items are merged into two aspects: themes and values. Themes are merged as ontology shows: superclass themes are covered by its subordinates, that is, those who are interested in "BMW" will also receive events with theme "car". Values are merged accordingly to basic arithmetic rules, that is, the subscription "length >5" will be covered by "length >3." Also, the next hops in route tables will be adjusted to adapt to these changes.

5 Experiment

With data from Douban, one of the China's most popular social networks, we experiment on people's interests to books and music. We do a fuzzy clustering to such interest information and classify users with similar interest into same clusters according to ontology library. We also find optimized cluster size and layers with this experiment. To evaluate data distribution efficiency, we build a pub/sub system on top of this network and compare average hop times, average transmission delay, and average transmissions per node with random network and hash clustering network. The network and pub/sub distribution scheme are simulated with NS3, a famous discreet simulator based on C++. Let us start with some performance indicator.

5.1 Performance Indicators

Major indicators of routing efficiency are average hop times, average transmission delay, average transmissions per node, and reception rate.

Average hop times (AHT): The average hop times for one event to transmit from publisher to subscriber.

Average transmission delay (ATD): The delay for every node to receive subscribed event, that is, time between events is published and received.

Average transmissions per node (ATN): Average number of events that nodes transmit during the experiment.

Reception rate: The ratio between number of events that one node receives and that of events that it interests/subscribes. In order to better evaluate the performance, our experiment is divided into three parts:

- a) To get the best from our system, we adjust the size of top-level cluster M and the number of levels L constantly to get the best network structure.
- b) We simulate the pub/sub process and compare AHT, ATD, and ATN under three conditions: first, random network; second, network leveled with hash functions; and third, network leveled with user information. Analysis of this result will tell to us to what extent is our system superior over others.
- c) To better simulate reality in social networks, we consider the situation in which certain number of random nodes lose efficacy or be replaced simultaneously. We randomly select nodes from current effective nodes and compare reception rate with different numbers of nodes being replaced or invalid.

5.2 Results

- a) When the top layer's size is large, event matching and routing are mainly within the same network. From its initialization to reception, event goes through relatively less hops and shorter transfer delay. However, in this case,

Fig. 3 Impact of clustering organization on data distribution

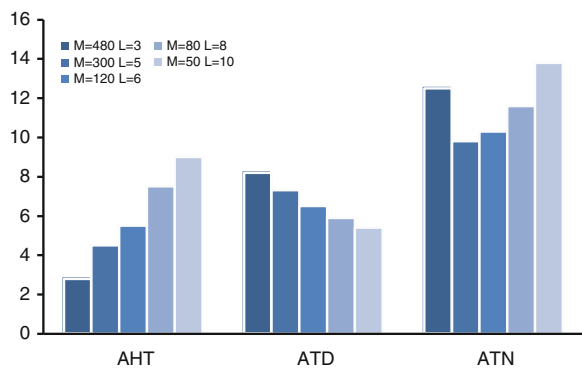


Fig. 4 AHT in different clustering networks

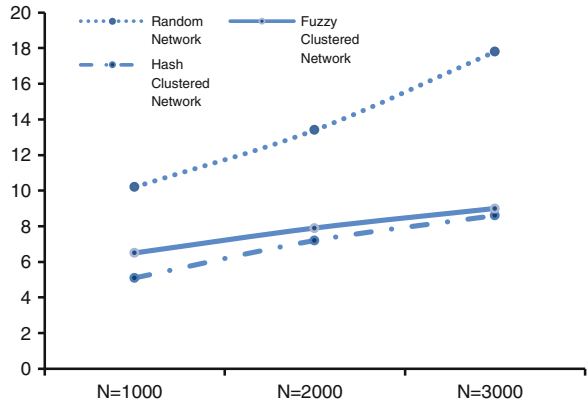


Fig. 5 ATD in different clustering networks

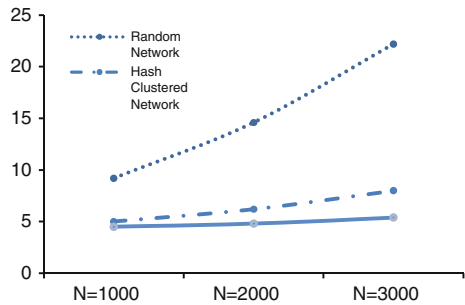
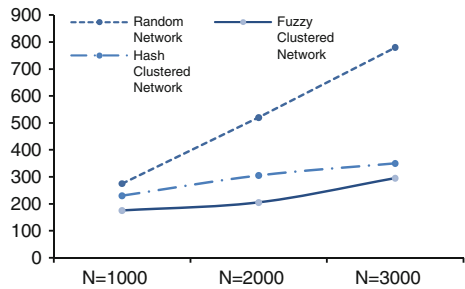


Fig. 6 ATN in different clustering networks



event matching may not succeed, wasting more time. With the increase in layers, event hopping is more and more done across clusters, the routing path is optimized, and average hops are reduced. Average number of events that each node receives diminishes with the increase in reception rate, while increases with more events routing across network clusters. Figure 3 shows how AHT, ATD, and ATN vary with different M and L (ATN data have been normalized).
b) Performance of different network structures is shown in Figs. 4, 5, 6 with AHT, ATD, and ATN. Compared with non-clustering network, clustering network has advantages in all three performance indicators. For average hop times,

Fig. 7 Average reception rate with replaced nodes

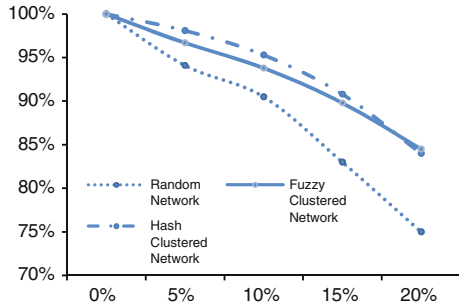
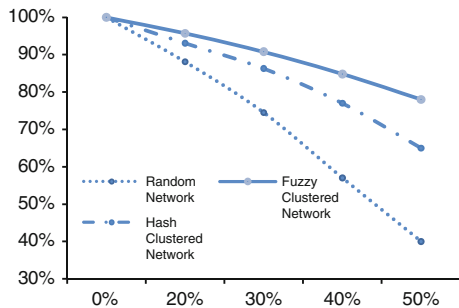


Fig. 8 Average reception rate with invalid nodes



fuzzy clustering network based on user information is a little weaker than hash clustering network because users’ interests are not that balanced. With the same network size, fuzzy clustering network has better load balance, requiring less network resources.

- c) When nodes become invalid or replaced, we publish events in the next ten time cycles and record average reception rate in the tenth cycle. As the graph shows in Figs. 7 and 8, even when large amount of nodes are invalid or replaced, fuzzy clustering network can still maintain more than 85 % reception rate. Since user information is copied and stored in more than one cluster, its performance is better than other networks when lots of nodes become invalid. We can then reach a conclusion that fuzzy clustering can adapt well to large-scale social network and remain robust even when the network is highly dynamic.

6 Conclusion

This paper proposes a fuzzy network clustering method and applies it in social network. We analyze users’ interest information and classify users with similar interests into same clusters and divide the entire network into a multi-layer topology. To evaluate its data distribution efficiency, we apply it with a theme-

based pub/sub system and experiment with 3,000 users' real data downloaded from Douban. Results show that fuzzy clustering network is superior to random network and hash clustering network in average hop times and transmission delay. Especially when large amount of nodes are replaced, this model can still maintain more than 85 % reception rate, which shows that it can not only adapt to highly dynamic environment but also remain robust. Future researches can focus on efficiency and accuracy of ontology library, improving load balance, or apply this methodology in more scenarios.

Acknowledgments This work is partially supported by China National Science Foundation (Granted Number 61073021, 61272438), Research Funds of Science and Technology Commission of Shanghai Municipality (Granted Number 11511500102, 12511502704), and Cross Research Fund of Biomedical Engineering of Shanghai Jiaotong University (YG2011MS38).

Appendix

Social network analysis (SNA) is a relatively new and still developing topic that focuses on the study of social relationship as a branch of the broader discipline named network analysis whose main object is to study the relationships between objects belonging to one or more universal sets. For better understanding of roles played by actors in social networks, different methods are invented to represent relationships among users. The so-called centrality indices have been introduced, where a member is viewed as central whenever she or he has a high number of connections with a high number of different co-members [1, 2]. This measure can be used to provide an ordering of the vertices in terms of their individual importance, but it does not provide any description of the way in which subsets of vertices influence the network as a whole. Another commonly used tool for representing social relationship, the adjacency matrix, is also limited to represent only pairwise adjacency, without shaping the strength of the relationship [3]. In [4], a technique to model multi-modal social networks as fuzzy social networks is proposed. The technique is based on k -modal fuzzy graphs determined using the union operation on fuzzy graphs and a new operator called consolidation operator. In [5], an approach to model uncertainty based on m -ary fuzzy adjacency relations and OWA operators is proposed. However, none of these methods incorporate fuzzy representation into data distribution of social networks nor do they take into account linguistically dynamics.

The dynamics of user's behavior traditionally pose great pressure to social media, especially in distributed models. For these applications, problems regarding privacy, extensibility, and single-point invalid toward centralized services make distributed service promising on the one hand, and user's highly dynamic behavior and huge themes make existing theme-based P2P methodology difficult to adapt on the other. Facing this situation, there come recently some data distribution models

especially for social network: non-structuralized topology-based model, structuralized topology-based model, and semantic topology-based model.

Social network topologies based on neighborhood relationship are non-structuralized topology. It is natural to use neighbors to build such topology. Maze [6] leverages social topology to organize nodes, while maintains node's status information and perform authorization with centralized servers. DSNF [7] allows users to store their information in credible servers, while neighbor information is stored by saving friends' URL. However, DSNF does not support off-line file reception. Moreover, social topology may lead to serious load imbalance, that is, nodes with more friends and neighbors have heavier loads. Once such nodes get off-line or invalid, and the connectivity of entire network will be seriously affected. Therefore, current, not much distributed social network services are built directly on social topology.

Major advantage of structuralized topology to non-structuralized topology is its support to highly efficient data routing. PeerSoN [8] uses third-party DHT (open DHT) as centralized servers. Users publish and update their status and information to DHT and get the latest information of their friends also by checking with DHT. For online friends, users can have direct communication with DHT, and for off-line friends, DHT acts as an intermediate transit point. SCOPE [9] and Safebook [10] also uses similar DHT to organize nodes, however, improve to some extent in load balance and privacy protection. Such methodology with third-party communication mechanism is not difficult to realize real-time data transfer and off-line file reception; however, dependency on third-party DHT may also lead to security and reliability problems like concentrated services. For other structuralized models, social network users' highly dynamic behavior makes maintenance a big problem.

Semantic topology based on users' interests is another way to organize nodes in social network. In GBDSNS [11], each node maintains two neighbor lists: one is neighbors with similar interest (semantic neighbor), and the other is random neighbors. Periodically it exchanges part of its list with another neighbor with same interest or randomly selected to update its own lists. It can realize off-line file reception in the way as real-time communication. However, its push communication tends to increase intermediate nodes' loads and leads to redundant communication.

In conclusion, existing data distribution application-faced social network cannot either adapt to highly dynamic social environment or support real-time and off-line communication in the same time. Therefore, more and more researches now focus on the improvement and evolution of these methods. PeerChatter [12] improves traditional non-structuralized topology and maintains a multi-layer random graph SkipCluster. Structured relationship between layers helps SkipCluster in highly efficient routing, while dynamically maintained random graph guarantee high robustness. However, PeerChatter still uses random routing mechanism, which can still improve, especially when it is used to match scarce subscriptions, and random routing always cannot realize expected data distribution efficiency. This paper is built on this. We combine non-structuralized topology

with fuzzy semantic definition, leaving most events transferred within the same network cluster. Data routing efficiency is enhanced a lot. Dynamically maintained user membership and information copies in multiple clusters also help to realize off-line file reception, while remain robust in highly dynamic environment.

References

1. Bonacich P (1987) Power and centrality: a family of measures *Am. J Sociol* 92:1170–1182
2. Borgatti SP (2005) Centrality and network flow. *Soc Networks* 27:55–71
3. Borgatti SP, Everett MG (1992) Regular block models of multiway, multimode matrices. *Soc Networks* 14:91–120
4. Nair, PS, Sarasamma ST (2007) Data mining through fuzzy social network analysis. In: *Proceedings of the 26th international conference of North American fuzzy information processing society*, San Diego, California pp 251–255
5. Brunelli M, Fedrizzi M, Fedrizzi M (2011) OWA-based fuzzy m-ary adjacency relations in social network analysis. Springer, Heidelberg
6. Hua C, Mao Y, Jianqiang H, et al. Maze (2004) A social peer-to-peer network. In: *Proceedings of the 2004 IEEE international conference on E-commerce technology for dynamic E-business*. Los Alamitos, USA, 290–293
7. Yeung CA, Liccardi I, Lu K et al (2009) Decentralization: the future of online social networking. In: *Proceedings of W3C workshop on the future of social networking*
8. Buchegger S, Schi D, Vu L et al (2009) Peerson: P2P social networking-early experiences and insights. In: *Proceedings of the second ACM Eurosys workshop on social network systems (SNS'09)*. New York, ACM: 46–52
9. Mani M, Nguyen A, Crespi N (2010) Scope: A prototype for spontaneous P2P social networking. In: *Proceedings of the 8th IEEE international conference on pervasive computing and communications workshops (PERCOM Workshop)*, pp 220–225
10. Cuttillo LA, Molva R, Strufe T (2009) Safe book: feasibility of transitive cooperation for privacy on a decentralized social network. In: *Proceedings of IEEE international symposium on world of wireless, mobile and multimedia networks (WoWMoW'09)*
11. Abbas A, Hales D, Pouwelse J et al (2009) A gossip-based distributed social networking system. Technical Report, PDS-2009-001, Delft University of Technology
12. Zheng Z (2011) High robust data distribution technology research faced to dynamic network environment. Doctor, National University of Defense Technology

Experimental Comparisons of Instances Set Reduction Algorithms

Yuelin Yu, Yangguang Liu, Bin Xu and Xiaoqi He

Abstract As techniques of data acquisition and data storage rapidly developed, more and larger datasets are very easily faced in machine learning. In order to avoid excessive storage and time consuming, and possibly to improve generalization accuracy by removing noise, several works presented as reduction techniques have been proposed. In this paper, firstly, we will review most traditional and typical reduction algorithms and find out their strengths and weaknesses, respectively. In addition, nine typical reduction algorithms are compared performing on 16 classification tasks. At last, some valuable directions for further research are proposed based on discussions and conclusion of traditional algorithms mentioned.

Keywords Instance-based learning · Nearest neighbor (NN) · Instance set reduction

1 Introduction

Researches about instances set reduction for large datasets have always been interesting and challenging in machine learning; especially, with the soaring development of data acquisition and data storage techniques at present, even larger datasets are easily and frequently confronted in machine learning. As a consequence, the complexity of most learning algorithms increases at least linearly with

Y. Yu · Y. Liu (✉) · B. Xu · X. He
Ningbo Institute of Technology, Zhejiang University,
Ningbo 315100, Zhejiang Province, China
e-mail: ygliu@acm.org

Y. Yu
College of Computer Science and Technology, Zhejiang University,
Hangzhou 310013, Zhejiang Province, China

the size of datasets. But, not all data points are necessary during training or other tasks of machine learning. What makes matter worse, some of them like noise could distort the results in machine learning application.

In order to avoid excessive storage and time consuming, and possibly to improve generalization accuracy by removing noise, several works presented as reduction techniques have been proposed. Most typical reduction algorithms could be roughly classified into the condensed nearest neighbor rule (CNN) serial [1–4] and instance-based learning (IBL) serial [5, 6]. From the beginning, P. E. Hart firstly proposed a nearest neighbor rule (NN Rule) reduction approach, and recently novel IBL with its variants were published, all of those previous works contribute a lot to the development of community of reduction techniques and its later property.

The rest of this paper is organized as follows. [Section 2](#) provides a review of typical reduction algorithms, such as the CNN, edited nearest neighbor (ENN) [3], IBL and decremental reduction optimization procedure (DROP) [6] serial and lists their advantages and disadvantages. [Section 3](#) presents experiments results to make comparisons with each other typical reduction techniques. By providing original training datasets substantial storage reduction while guaranteeing their generalization accuracy, each algorithm shows their outperformance in storage reduction or generalization accuracy maintaining. The final section provides discussions about those experiments, and some remarks and conclusions are drawn.

2 Instances Set Reduction Algorithms

A lot of researches have addressed the problem of training set reduction. In this section, we will review several traditional and typical reduction algorithms, and from their algorithm designing, seek to find out their strengths and weaknesses, respectively.

2.1 Nearest Neighbor Rules

2.1.1 Condensed Nearest Neighbor Rule

P. E. Hart was the first to propose an approach, namely the “CNN” [1], to reduce the size of stored training dataset for the task of machine learning. The basic idea of his algorithm is to select a subset S from original training dataset T by eliminating very similar patterns and those that do not add additional information. At the same time, S should be as effective as T when classifying unlabeled data points in later tests. Instances from T holding each class are randomly selected and put in S . Scan through T to make sure very instance could be correctly classified by instances from S . An instance, which has been misclassified, would be added to S .

This process is iterative until no instances from T are misclassified. F. Angiulli extended it in the “fast condensed nearest neighbor rule” (FCNN) [2], creatively created subsets serving as consistent training sets instead of selecting from T .

However, as the seeds are randomly selected for S , the CNN along with the FCNN is very sensitive to noise, which can decrease generalization accuracy and hinder storage reduction correspondingly in the procedure of machine learning when noise is retained rather than deleted.

2.1.2 Edited Nearest Neighbor Rule

To address the noise problem, Wilson has developed the “ENN” [3] algorithm, in which S starts out the same as T , and then each instance in S would be removed, if it does not agree with the majority of its k nearest neighbors. This edits out noisy instances as well as close border points, leaving smoother decision boundaries. On the other hand, it retains all points, locating relatively far from boundaries, i.e., internal points, which, however, keeps the ENN from reducing the storage requirements as much as most other reduction algorithms.

Tomek extends the ENN [6, 7] with his own All-kNN [8] method of editing, assigning $i = 1$ to k to the kNN algorithm and removing any instance from S , if it has been misclassified by most of those kNN tests. Strictly speaking, this algorithm serves more as a noise filter than a regular reduction algorithm.

Many other creative related works like selective nearest neighbor rule (SNN) [9], reduced nearest neighbor rule (RNN) [10], repeated edited nearest neighbor rule (RENN), etc., all contribute a lot to the development of the CNN serial reduction algorithms.

2.2 Instance-Based Learning algorithms

Besides the CNN serial algorithms, novel IBL [11–13] algorithms have always been proposed in recent literature. Aha et al presented a series of IBL algorithms. The basic and simple IBL algorithm is IB1 (instance-based learning algorithm 1) [5, 14], which is simply the 1-NN algorithm, and has been extended as a baseline.

2.2.1 IB2 and IB3

Later IB24 is quite like the CNN except that S is initially empty and that IB2 does not pass through training set T repeatedly, but those improvements cannot help IB2 break the curse of noisy points.

As a result, IB3 [5, 14] is developed aiming to address IB2’s problem by retaining only acceptable misclassified instances. An instance is acceptable, only if

it has a significantly higher confidence on accuracy statistically than its class' frequency.

Due to its reduced sensitivity to noise, IB3 can achieve greater reduction in storage and also higher accuracy than IB2 statistically.

2.3 *Decremental Reduction Optimization Procedure*

Wilson et al. [6] suggest a series of the DROP for sets reduction based on the kNN algorithm where each algorithm improves the previous one. DROP1–5 combines the basis of RNN [15] and other delicately designed rules and preprocessing, which achieves quite great performance in experiments, especially the DROP3 embedded with a noise filter.

DROP1 is based on the following rule: an instance x is removed only if at least some of its neighbors from the same class in S can be classified correctly without x . As we all know, noisy instances are surrounded with instances holding different classes. So it would cause problems when some noisy instances are a portion of the original training set. DROP2 tries to address this problem by removing an instance on all instances of initial training set T rather than S .

But, when noisy instances locate on or quite close to decision boundaries, DROP2's removal rule would change. It would eliminate border points, which are very important in deciding border and should be retained. Therefore, in DROP3, a noise filtering is applied to distinguish noise from border points. However, sometimes DROP3 removes an overly large number of instances.

DROP4 improves DROP3 with more carefully filtering noise, an instance is removed only if it is misclassified by its k nearest neighbors, and its removal does not affect classification of other instances.

DROP5 modifies DROP2 in removal rule that instances are considered for removal if they are nearest to their nearest neighbors holding different class.

Some other meaningful researches on prototypes [10, 16] have been carried out before. For example, Chang introduces his own algorithm by treating each instance in as a prototype. The nearest two instances holding the same label are merged into a single prototype, e.g., centroid. This process is repeated until classification accuracy started to suffer. It turns out to achieve quite good results.

From what we discussed above, the problem of all those reduction algorithms can be formulated as seeking for a representative subset of instances. Algorithms' storage reduction ability mostly depends on the way they are dealing with similar points, which are redundant for later tasks of machine learning. How to handle noisy points, including close border points, greatly decides algorithms' maintaining generalization ability. If an algorithm combines prototype theory, which is quite effective in similar points, reduction, with noise filter, would achieve great performance in both generalization accuracy and storage reduction.

3 Experimental Results

In the machine learning community, there have been much previous works on the training set reduction formulated as the problem of deciding which instances to store before generalization. On the one hand, Hart and other researchers, such as Ritter, Gates, Wilson et al., attempt to reduce the size of the training set with the nearest neighbor rule, which directly result in CNN serial algorithms' springing up and their prosperity. On the other hand, the IBL algorithms demonstrate their great performance in training set reduction treating instances from training set as prototypes, especially when noisy instances cover a portion of original training set.

3.1 Setup for Experiments

The datasets used in this paper are described and explained as follows. Most of them are downloaded from the UC Irvine Machine Learning Repository, and they can widely cover the physics, computer science, life science, and business fields (details are given in Table 1). Note that some of the datasets contain several sub-datasets, for example, wine quality actually includes wine quality (red) and wine quality (white) datasets. This also applies to the cardiocography and multiple features datasets. As a result, a total of 16 sets are examined in this paper.

The tenfold cross-validation is applied for each experiment. More precisely, each dataset is divided into ten subsets. Each one of them is used in turn for performance estimation as the test set. The remaining subsets merge to form the training set. Based on several trial runs for each dataset with each reduction technique, the average accuracy and storage requirements are reported. kNN classifier (assuming $k=7$) serves for verification of performance. Regarding the similarity function, we choose Euclidean distance function after the standard deviation process. Concerning complex circumstances like the heterogeneous features in future application, the heterogeneous value difference metric (HVDM) [17] would be a more suitable option.

Table 1 Datasets used in experiments

Dataset name	Size	Feature	Classes
Cardiocography	2,126	23	10 or 3
Image segmentation	2,310	19	7
MAGIC gamma telescope	1,9020	11	2
Spambase	4,601	57	2
Steel plates faults	1,941	27	7
Wine quality	4,898	12	11
Page blocks	5,473	10	5
Statlog (satellite)	6,435	36	6
Musk	6,598	166	2
Multiple features	2,000	649	10

3.2 Comparisons Between Reduction Algorithms

In this subsection, the traditional instances set reduction algorithms, such as CNN, ENN, IB2, IB3, and DROP1–5, which cover both the CNN serial and IBL serial, would perform on the same datasets, and comparisons between each other are made.

From Tables 2, 3, we could safely draw the conclusion: with a certain proportion of original instances removed, most reduction techniques have maintained their generalization accuracy with varying degrees of injury, some of which could be acceptable anyway. Meanwhile, each reduction algorithm’s design promises their pretty good performance on different datasets, some even improve generalization ability, such as ENN on Mfeat-zer, Drop3 on MAGIC, Drop4 on wine quality (red). Figure 1 shows us the count statistics datasets each reduction algorithm achieved pretty good performance. Note that “Good performance” case here means one kind of reduction technique run on a reduced dataset achieves accuracy dropping around 1.0

To evaluate the performance of a training set reduction algorithm, there are various criterions that can be used such as storage reduction, generalization accuracy (achieved by test classifier), noise tolerance, time requirement. But among them, storage reduction and generalization accuracy are main goals of reduction algorithms. The performance can be roughly evaluated if we just take the most main two goals into consideration. It is known to us that the performance has

Table 2 Experimental results on 16 data sets

Data Sets	kNN	%	CNN	%	ENN	%	IB2	%	IB3	%
CTG (3C)	92.11	100	88.26	14.69	90.13	91.09	87.30	14.69	84.42	10.31
CTG (10C)	74.79	100	69.58	38.51	67.87	72.88	70.54	38.09	69.53	33.13
Image	84.11	100	74.11	29.31	81.99	83.17	74.20	29.22	79.52	17.33
MAGIC	84.53	100	77.90	23.19	84.13	85.01	77.90	23.19	80.95	3.59
Faults	74.49	100	67.39	36.41	70.58	74.84	68.06	35.99	70.22	23.96
Wine quality (red)	58.63	100	56.22	49.45	59.16	63.87	55.72	47.88	54.91	24.98
Wine quality (white)	56.88	100	50.51	52.41	55.10	59.14	51.23	52.58	50.59	24.80
Page blocks	96.83	100	95.60	5.72	95.92	96.67	95.40	5.69	91.95	2.93
Satellite	56.21	100	47.21	52.37	57.16	56.80	47.27	52.36	48.24	44.80
Musk	97.07	100	92.71	12.09	95.70	96.48	92.71	12.09	88.49	6.61
Mfeat-fac	96.92	100	91.85	16.62	95.50	96.02	91.20	16.58	92.05	21.78
Mfeat-fou	76.61	100	72.25	36.76	65.45	70.43	72.55	36.36	73.00	37.34
Mfeat-kar	94.55	100	89.30	17.91	91.95	93.09	89.25	18.24	91.20	20.04
Mfeat-mor	70.00	100	62.26	38.48	70.53	72.47	62.91	38.95	65.50	30.66
Mfeat-zer	81.13	100	76.75	28.37	82.60	84.31	77.25	28.33	78.90	26.53
Spambase	89.46	100	83.76	20.44	86.87	88.29	83.76	20.44	82.28	10.51
Average	80.27	100	74.73	29.55	78.17	80.29	74.83	29.42	75.11	21.21

The left column presents accuracy, and the right shows the percent of original training set retained by reduction algorithm

Accuracy and storage percentage for CNN, ENN, IB2, IB3

Table 3 Accuracy and storage percentage for Drop1–Drop5

Data sets	Drop1	%	Drop2	%	Drop3	%	Drop4	%	Drop5	%
CTG (3C)	85.96	11.78	90.87	17.60	89.17	11.90	90.71	15.66	90.29	14.86
CTG (10C)	69.63	31.80	73.16	42.22	69.05	29.63	71.18	39.59	72.25	34.32
Image	74.42	14.23	81.99	21.40	82.12	15.18	83.20	18.28	83.25	17.27
MAGIC	82.03	12.88	84.61	21.96	85.06	10.62	84.94	15.23	84.32	14.88
Faults	65.02	21.68	71.71	30.05	70.01	20.03	71.10	25.99	70.63	24.75
Wine quality (red)	56.85	31.25	60.28	38.44	59.16	25.74	60.66	33.59	60.79	32.06
Wine quality (white)	52.88	33.06	55.94	42.87	56.37	23.94	56.29	35.68	56.62	34.80
Page blocks	93.04	4.04	96.51	5.63	95.98	4.05	96.51	5.10	96.40	4.75
Statellite	53.47	30.91	54.42	40.04	56.75	17.32	55.42	31.05	54.14	31.20
Musk	88.63	11.81	96.21	18.93	94.85	13.35	95.39	15.84	94.76	16.93
Mfeat-fac	87.70	13.28	94.00	24.09	93.90	23.46	93.55	24.08	91.55	18.93
Mfeat-fou	68.90	24.44	71.90	31.80	66.80	23.06	71.25	30.86	74.25	27.77
Mfeat-kar	86.50	14.81	92.40	25.01	91.50	23.61	92.25	24.75	90.05	19.64
Mfeat-mor	68.15	19.43	70.05	27.51	69.70	13.94	68.46	19.89	69.34	19.53
Mfeat-zer	76.20	19.28	79.95	28.54	80.10	22.42	80.00	25.82	79.40	22.39
Spambase	81.85	11.94	87.56	18.63	86.79	14.26	87.38	16.83	87.11	15.88
Average	74.45	19.16	78.85	27.17	77.96	18.28	78.64	23.64	78.45	21.87

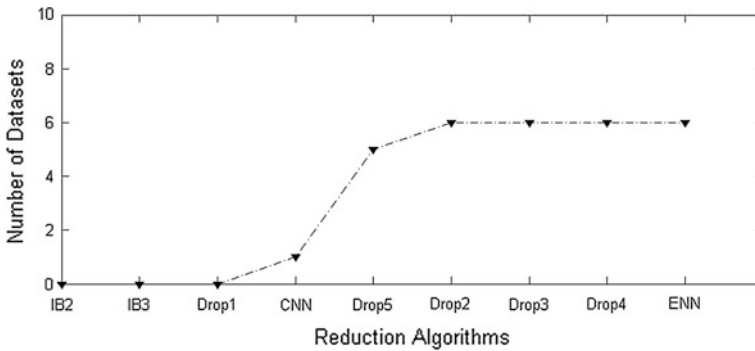


Fig. 1 Good performance datasets count and statistics

an obviously positive correlation with generalization accuracy (the bigger the better) and a negative one with the size of set output (the smaller the better). To interpret it explicitly, we give the following function:

$$pe_i = \frac{acc_i(1 - s_i)}{acc_0}, \text{ where } 0 < s_i, acc_i < 1.0$$

From Fig. 2, we could learn that Drop3–5 which have more carefully dealing with noise and decision boundaries outperform others in comprehensive consideration with generalization accuracy and storage reduction. Just as we mentioned above, the way you weight data points to find out similar points and noise including close noise is very important in reduction approaches.

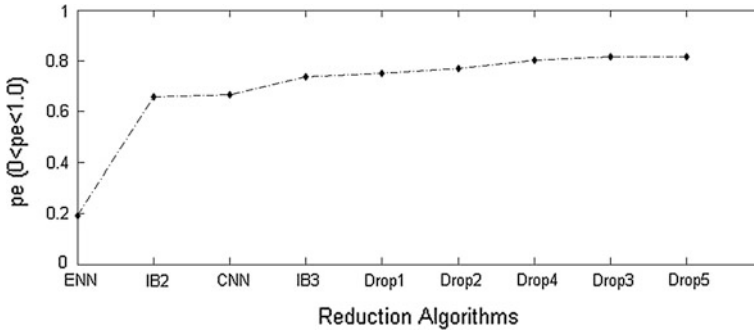


Fig. 2 pe of each reduction algorithm

In conclusion, Drop2–5 and ENN have shown their great performance in maintaining generalization accuracy in our experiments, trying to keep the completeness of original training set; some of them even improve several reduced subsets' generalization ability by eliminating or reducing noisy points. By taking pe into consideration, Drop2–5 algorithms demonstrate their comprehensive performance in both maintaining generalization accuracy and storage reduction. However, ENN tries to reduce noise while retains all points, locating relatively far from boundaries, i.e. internal points, which keeps ENN from reducing the storage requirements as much as other reduction algorithms.

4 Discussions and Conclusions

In this paper, we have firstly reviewed several traditional and typical reduction techniques. With discussion of existing reduction algorithms, we provided their advantages and disadvantages analysis based on previous works and our own thoughts about reduction approaches' designing. By treating each instance as a prototype, the problem of redundant points' reduction would be solved by prototype merging. And noisy point should be handled carefully because of its underlying distortion of generalization ability and storage reduction. Note that each technique is very effective in a special area and under some special circumstances.

Acknowledgements This work is supported by Ningbo Natural Science Foundation of China (Grant No. 2009A610083, 2011A610177) and partially supported by Zhejiang Provincial Natural Science Foundation of China(Grant No. Y1101202).

References

1. Hart P (1968) The condensed nearest neighbor rule (corresp.). *Inf Theory IEEE Trans* 14:515–516
2. Angiulli F (2005) Fast condensed nearest neighbor rule. In: *Proceeding of 22nd international conference on, machine learning*, pp 25–32
3. Wilson DL (1972) Asymptotic properties of nearest neighbor rules using edited data. *IEEE Trans Syst Man Cybern* 2:408–421
4. Cover T, Hart P (1967) Nearest neighbor pattern classification. *Inf Theory IEEE Trans* 13:21–27
5. Aha DW, Kibler D, Albert MK (1991) Instance-based learning algorithms. *Mach Learn* 6:37–66
6. Wilson DR, Martinez TR (2000) Reduction techniques for instance-based learning algorithms. *Mach Learn* 38:257–286
7. Tomek I (1976) An experiment with the edited nearest-neighbor rule. *IEEE Trans Syst Man Cybern SMC* 6:448–452
8. Kubat M, Matwin S (1997) Addressing the curse of imbalanced training sets: one-sided selection. In: *ICML*, pp 179–186
9. Ritter GL, Woodruff HB (1976) S.R.L.T.L.I.: an algorithm for a selective nearest neighbor decision rule. *IEEE Trans Inf Theory* 21(6):665–669
10. Domingos P (1995) Rule induction and instance-based learning a unified approach. In: *Proceedings of the 14th international joint conference on artificial intelligence*, vol 2. *IJCAI'95*, San Francisco, CA, USA, Morgan Kaufmann Publishers Inc, pp 1226–1232
11. Cameron-Jones R (1995) Instance selection by encoding length heuristic with random mutation hill climbing. In: *Eighth Australian joint conference on artificial intelligence*, Canberra, pp 99–106
12. Skalak DB (1994) Prototype and feature selection by sampling and random mutation hill climbing algorithms. In: Cohen WW, Hirsh H (eds) *ICML*, Morgan Kaufmann, pp 293–301.
13. Gonzalez C, Dutt V (2011) Instance-based learning: integrating sampling and repeated decisions from experience. *Psychol Rev* 118:523–551
14. Aha DW (1992) Tolerating noisy, irrelevant and novel attribute in instance-based learning algorithms. *Int J Man-Mach Stud* 36:267–287
15. Gates GW (1972) The reduced nearest neighbor rule. *IEEE Trans Inf Theory* 18:431–433
16. Chang CL (1974) Finding prototypes for nearest neighbor classifiers. *IEEE Trans Comput* 23:1179–1184
17. Segata N, Blanzieri E, Delany S, Cunningham P (2010) Noise reduction for instance-based learning with a local maximal margin approach. *J Intell Inf Syst* 35:301–331

Recycla.me: Technologies for Domestic Recycling

Guadalupe Miñana, Victoria Lopez and Juan Tejada

Abstract Mobile applications have achieved a great development in the last years. Some of the existing tools are useful for academics and industry. In this regard, we propose in this paper a new mobile application, called Recycla.me, to help of social awareness about proper recycling of everyday materials. This is a mobile application developed under the Android OS whose main objective is to help children between 7 and 11 years to learn about this topic. The application works on any type of product that we can find in a supermarket, and it is based on reading the bar code that usually has the product. In this way, by reading the bar code, the application will guide step-by-step in the proper mode of recycling of each component in order to make a good recycle of the materials. In addition, the application works when the product has not got a bar code. In this case, the user is guided through a simple questionnaire to identify all the materials the product has. Finally, the application shows the appropriate recycling bin where each item must go. The application uses some operational research techniques.

Keywords Mobile applications · Android · Recycling · Bar code · Sustainable world · Operational research techniques

G. Miñana (✉) · V. Lopez
Mobile Technologies and Biotechnology Research Group (G-TeC),
Informatics Faculty, Complutense University, Madrid, Spain
e-mail: guamiro@fdi.ucm.es

V. Lopez
e-mail: vlopez@fdi.ucm.es

J. Tejada
Mobile Technologies and Biotechnology Research Group (G-TeC),
Mathematics Faculty, Complutense University, Madrid, Spain
e-mail: jtejada@mat.ucm.es

1 Introduction

Nowadays, governments and population are increasingly aware of the need to recycle household waste properly. There is a lot of literature already published in this issue (see [1] for example). However, the information comes to the people by different ways and this information is not always reliable. Many people have doubts about how to recycle a product either because they do not know to identify the component materials or because they do not know about the appropriate recycling bin. Generally, these people do what a neighbor or friend do.

One way to address this problem is to teach children to recycle properly. There is evidence that people learn more than 70 % of their knowledge and habits during childhood, and this is why we have develop our first tool about recycling for children recycling habits. An attractive way to teach children to recycle is using new technologies. The work presents in this paper is focused on this idea. We have developed a new mobile application (called Recycla.me) which is designed for 7–11-year-old children. They learn how to recycle as a habit by using the application. This application allows, by capturing the bar code of the product to recycle, to know in real time the proper recycling of each of the materials that make up the product. If the product does not have a bar code, the application allows the user to identify its components through icons. Once identified, the application shows the user the appropriate recycling bin where they should throw each one of the items or materials.

The paper is organized as follows. [Section 2](#) describes the State of the Art. The tool architecture and functionalities are specified in [Sect. 3](#). [Section 4](#) shows how our mobile application works and some results. And finally, in [Sect. 5](#), we present conclusions and future work.

2 State of the Art

In the market of mobile applications, we can find a lot of applications in the same line of Recycla.me. Some of them are described as follows.

2.1 *Where to Recycle*

This application works as a complement to a web site [2]. This application provides information on where to find recycling points in the different cities of Spain. When the user has selected one of them, on screen appears information about the type of materials that are recycled on it, schedules, and so. In the other hand, this application allows the user to include new recycling points to the map provided by the application (Google maps). This application does not have an educational focus, because neither shows how to recycle a product nor gives tips on recycling.

2.2 Recycling Guide

Developed by a Spanish non-governmental organization (Ecoembes), this application displays a list of material that can be recycled. Once the user has selected one among the possibilities, the application displays the appropriate recycling bin where to dump it. In the other hand, it also allows the user to select a specific recycling bin. In this case, the application displays information about the materials which can be thrown on it. As shortcomings, this tool does not show location of the recycling point that there is in the city, neither displays tips on recycling.

2.3 Recycling Heroes

This application has an interface more attractive than the applications described above and allows the following actions:

- When the user selects a material, the application displays the appropriate recycling bin where you should throw it.
- The application displays the user where is the nearest recycling point.
- It shows statistics on natural resources that are benefit when a certain material is recycled.

This application is more complete than the other applications, but it has some drawbacks: it is only available for iOS and it can be only use in Valencia, a very nice city in Spain.

3 The Architecture and Its Functionalities

3.1 Software Architecture

The structure of Recycla.me software is based on the data flow between the interface point and the database. The user introduces input data by means of queries on the device screen. There are two ways to feed the database. In both cases, a bar code is read. If the bar code is already stored in the database, the application simply searches for information related and it is displayed on the screen. There are two ways to feed the database. In both cases, a bar code is read. If the bar code is already stored in the database, the application simply searches for information stored and it is displayed on the screen. Otherwise, the application stores the bar code together with a set of information that gets from the user and a decision tree. This procedure is done by means of a learning algorithm based on decision tree techniques [3, 4].

The decision tree is used for items with more than one material, or if there is any doubt in the previous step. After an initial design, the tree has been refined by experts in recycling by means of linguistic labels, in order to maximize the knowledge to implement and to correct errors. Thus, a structure is obtained with eight initial nodes that determine the application that will travel path.

Of those eight nodes are chosen those whose branches you want to explore, and then run through one after another, selecting and providing the information necessary to enable the system to provide an adequate response.

The result is a decision tree that introduces some intelligence to the system in running time. All information about this software develop can be found in the research web site [5]. Figure 1 shows the view of the Java activities that compound the project.

3.2 Functionalities

Recycla.me, the application that we have developed, provides interesting improvements: On the one hand, our application, using the bar code of the product to recycle, shows the user the different materials that it has and the recycling bin where they should throw each one of them. This is the main improvement of our application because in the other tools, the user can select materials but not products, and this implies that he need to know in advance the materials that have the product.

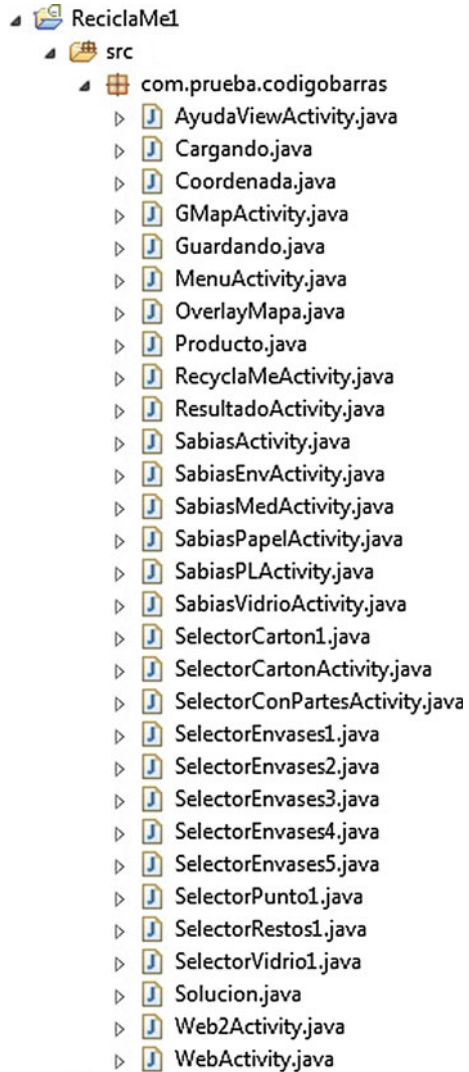
On the other hand, it allows the user to know where is the nearest recycling point. In order to find the optimal route between the user's position and the different recycling points in the city, our application uses operational research techniques. Also, for the children to learn the importance on recycling, the application shows some tips, curiosities, and reminders about the recycling of different materials.

Finally, Recycla.me has an attractive and intuitive interface that has been designed mainly for children. To facilitate children's learning, the interface always associates the color of the icons representing the material to be recycled to the actual color of the recycling bins. This part of the project has been developed in close collaboration with the Department of Environment of the City of Madrid. The Recycla.me mobile application offers to the user a great variety of functional requirements. Figure 2 shows the initial screen of the application. In it, there are two icons that allow the user to choose between the two options described as follows.

- **With bar code**

The user can get information about how to recycle a product by two ways: one of them is by scanning (capturing) the bar code of the product. Once captured the bar code, the application checks if it is in its database. If the bar code is already

Fig. 1 Activities Java view



stored, the application shows the user each one of the product’s materials and the recycling bin where they should throw each one.

• **Without bar code**

When a bar code is not available or it is not already stored in the database, the user has to select this option. It works as follows:

If the product has one only material, (as cans of soft drink), the application allows the user to identify the product through pictures. As seen in Fig. 3, the screen displays an icon by each kind material that can be recycled. When the user

Fig. 2 Initial screen of Recycla.me



press one of them, the application shows the user the recycling bin where they should throw this product.

To the products that have more than one material (as biscuit boxes), the application uses a tree decision. We have used this technique of operational research to efficiently optimize the decisions that have a large number of options.

For this kind of products, the user has to press the icon more than one material (column three, row four in the Fig. 3); then, the user is guided through a simple questionnaire to identify all the materials that has the product. Once identified, the

Fig. 3 Material classification



application shows the user the appropriate recycling bin where they should throw each of the materials.

- **Product verification**

Verification by an adult person is needed when bar codes are not available and after the questionnaire is filled in. As the application is targeted for children, the application warns the children to ask an adult for approval on the solution obtained.

- **Recycling tips**

These elements show up when the application informs the user where they should throw the material. They are interactive to make the application more interesting to the children. There are three types:

- *Remember* this element contains important information to consider when recycling.
- *Eye* this is an interactive element that displays advices on recycling. Whenever the icon is pressed, a random tip on recycling pops up.
- *Know that?* Is another interactive element that whenever it is pressed displays a series of figures or statistics on natural resources that are benefit when a certain material is recycled.

Figure 4 shows an example of these elements. In this case, the material to recycle is a book. The application informs the user that this product should be thrown in the blue bin. In this screen appear the elements Remember, Eye, and Did you know...?. The figure the middle shows the result of press the eye, and the figure of the right shows the result of pressing Did you know...?.

- **Viewing a recycling point on the map**

When the application informs the user about throwing the product away in a recycling point, this icon allows finding the nearest recycling point.

- **Help**

By touching this icon, the application informs the user how it works.

Apart from the previous requirements, related to the application functionality, there are others related to evaluate the tool quality from a technical point of view. In this regard, we expect that our application will be scalable, with high performance, very intuitive to use and small. These objectives seem to be achieved from the obtained results of the tests.

Recycla.me has been developed for the Android operating system [6, 7] in the Java programming language [1] with Eclipse. We have developed a database that stores information about the bar code of a product, the materials that it has, and the recycling bin where they should be thrown each one.

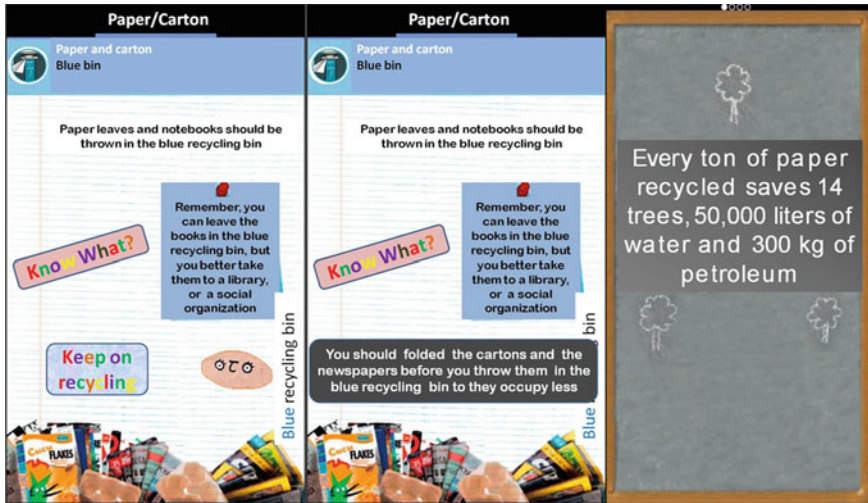


Fig. 4 Recycling tips

At present, the application Recycla.me can be downloaded from the web site of the research team G-TeC [5]. To use, Recycla.me is necessary to have a mobile device with the following features: Android OS 2.2 and up, camera and data transfer via WiFi or 3G. Once it is installed in our mobile for executing the application, it is enough to click on its icon.

4 Results

Recycla.me has been tested in several primary schools in Madrid, Spain, during the month of May 2012. These sessions were driven by the government employees and they consisted of the following:

- A conference about the importance of recycling
- A demonstration about how to use the application Recycla.me
- A game to test Recycla.me
- A test to evaluate the application.

The recycling bins were simulated with colored cartons. These bins were placed in the room, and a product to recycle was given to each student. Under the supervision of an adult, the students used the application to know the proper recycling of each of the materials that make up its product. The children were in age from 8 to 9 years.

After the session, students filled out a test to evaluate the application. The results were very positive; the students think that Recycla.me is very friendly and easy to use. Here are highlights of what more liked to the children:

- The interface is attractive and intuitive.
- The advice about recycling is very interesting, and most of the children acknowledge that they have learned a lot.
- Work with the bar code.

5 Conclusions and Future Work

We have developed a mobile application that makes easier the learning of the proper recycling of any product. The main feature of our application is that the user does not need to know in advance the materials the product has, because Recycla.me guides the user through a simple questionnaire to identify all its materials. Once identified, the application shows the user the appropriate recycling bin where they should throw each of the materials. Also, the application teaches the children tips, curiosities, and reminders about the recycling of different materials.

Recycla.me has been tested in the sessions about recycling that each year organizes the city of Madrid during the month of May in the schools. Students rated the application with an average grade of 8 into the range 0–10.

The developed application can be extended in several ways: (1) adapt the application for the disabled, for example, by adding voice as descriptive resource; (2) extend the application for other operating systems such as iOS (iPhone) or Windows Mobile, in order to broaden the number of user; (3) develop a new version for adults; and (4) connect the application with social networks as Facebook or Twitter.

References

1. Arnold K, Gosling J, Holmes D (2005) Java programming language (java series), 4th edn. Addison Wesley Professional
2. <http://dondereciclar.com>
3. Deng H, Runger G, Tuv E (2011) Bias of importance measures for multi-valued attributes and solutions. In: Proceedings of the 21st international conference on artificial neural networks
4. Yuan Y, Shaw MJ (1995) Induction of fuzzy decision trees. *Fuzzy Set Syst* 69:125–139
5. www.tecnologiaucm.es
6. Haseman C (2008) Android essentials (books for professionals by professionals) 1st edn. First Press
7. Gargenta M, Learning android, 1st edn. O'Really Media, Inc
8. Jefferson B et al (2000) Technologies for domestic wastewater recycling. *Urban water*, Elsevier Science, pp. 285–292

Caching Mechanism in Publish/Subscribe Network

Hongjie Tan, Jian Cao and Minglu Li

Abstract Content-based Publish/Subscribe network is a flexible communication model. It can support communication using the content of message instead of the network address, which will meet the need of data transmission in large scale. In traditional Publish/Subscribe network, messages are not stored in the network and subscribers can only receive the messages published while they are online. However, in some dynamic scenes where the users join and leave the system dynamically, a new user might be interested in the messages published in the past. This paper proposes a distributed caching algorithm to store messages and support subscribing to historical messages in Publish/Subscribe network, while maintaining the loosely coupled and asynchronous communication of the network. By comparing with the other two caching algorithms, the proposed caching algorithm outstands in persistence capacity, overhead of history retrieval, user response delay, and scalability.

Keywords Publish/Subscribe · Distributed caching · Message retrieval · Mobility support · Data persistence · Historical data · Content-based networks

H. Tan (✉) · J. Cao · M. Li
Department of Computer Science and Engineering, Shanghai Jiao Tong University,
Shanghai, China
e-mail: luckythj@sjtu.edu.cn

J. Cao
e-mail: cao-jian@cs.sjtu.edu.cn

M. Li
e-mail: li-ml@cs.sjtu.edu.cn

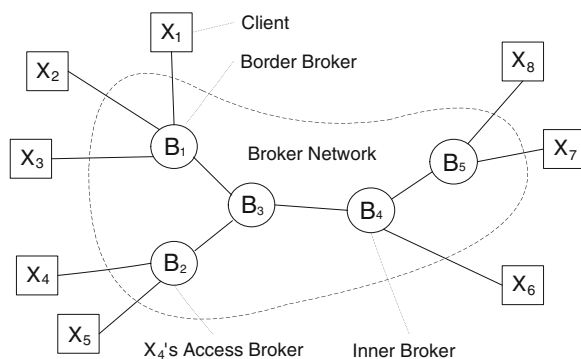
1 Introduction

With the wide use of Internet technology and rapid development of mobile computing, pervasive computing, as well as the Internet of Things technology, a distributed system may contain thousands of nodes that may be distributed across different geographic locations and have different behaviors. These limitations make distributed systems need a more flexible communication model to adapt to the dynamic and scalability. Publish/Subscribe paradigm, with its loosely coupled and asynchronous communication characteristics, is becoming an important model in designing distributed systems and is considered to be one of the most promising network architectures to solve many challenges of Internet today.

As shown in Fig. 1, Publish/Subscribe system consists of a set of *Clients* and a *Notification Service*, which exists between the clients to decouple their communication. The *Clients* can be divided into *Subscribers* and *Publishers*. Subscribers submit their *Interests* to the *Notification Service*, while *Publishers* deliver *Messages* (or called *Events*) to the *Notification Service*. The *Subscribers* will receive the published messages meeting their *Interests* through the *Notification Service*. The Publish/Subscribe network composed of *Brokers* playing the role of *Notification Service*, and it collects *Subscriptions* and routes the published *Messages* to the appropriate *Subscribers* [1]. Publishers and subscribers do not need to know each other's address information, which provides decoupling in communication address [2]. There are many existing message notification services using Publish/Subscribe systems, such as Gryphon [3], Siena [4, 5], JEDI [6], Rebeca [7], Hermes [8], and Elvin [9]. Existing studies of Publish/Subscribe system mainly focus on the message matching, routing algorithms, system scalability, and security features.

In Publish/Subscribe system, it is ensured that each available message is delivered to subscribers who subscribed to that message (completeness property) [10]. However, in traditional Publish/Subscribe system, clients must be online during the communication; otherwise, the completeness property will not be satisfied.

Fig. 1 Basic Publish/Subscribe network



With the popularity of mobile computing and the increasing requirement for quality of service, the assumption that the communicating clients are always online will not apply to many scenarios. For example, when the client is a mobile device, the Publish/Subscribe system must be able to support the client's mobility. When the mobile client joins the system dynamically, it is likely to be interested in the messages generated in the past. However, the traditional Publish/Subscribe system cannot meet this scenario. Therefore, persisting messages and retrieving historical messages is a challenging and significant topic in Publish/Subscribe system. The caching mechanism described in this paper is a solution to this problem. It persists messages in the Publish/Subscribe network and provides a quick and reliable access to the historical messages.

This paper proposes a caching algorithm for message persistence and extends traditional Publish/Subscribe system to enable historical message retrieving. Finally, we evaluate the proposed caching algorithm from the aspects of persistence capacity, overhead of history retrieval, user response delay, and scalability.

2 Related Works

In content delivery network (CDN), there are many researches on caching algorithm. Pierre and van Steen [11] propose a collaborative CDN called Globule, which is composed of servers distributed in various geographic locations. These servers collaborate and make use of each other's resources to improve the entire quality of service. The main purpose of the CDN is how to deploy the caches of content to obtain improvement in performance. So, caching algorithms in CDN are mainly focused on how to place the content caches and make them close to users.

These traditional caching algorithms are mainly focused on the replicating information to improve the quality of service of data access and reduce network overheads. The replica of the information is usually stored close to users physically or logically. However, caching algorithm used in this paper mainly focuses on persisting information. In the field of Publish/Subscribe system, data persistence has not been the research focus. But there still have been some academic achievements.

Cheung et al. [12] and Singh et al. [13] propose a Publish/Subscribe system supporting historical data recovery. They use databases as predefined caching points. Each database is attached to a filter so as to store the specified part of the messages. In these systems, we need to add additional databases to support message persistence. However, when the Publish/Subscribe system scales up, the use of centralized database may bring many negative effects, such as increased equipment costs, performance bottleneck caused by accessing database, and single-point failure. In this paper, we make use of the inner brokers to form a distributed storage system, so as to adapt to the large-scale scenarios and reduce unnecessary costs.

Sourlas et al. [14] propose a caching algorithm, which stores the messages in specified brokers, and provide a retrieving algorithm, which can locate the caching points easily without additional protocols. Diallo et al. [15] introduce a caching algorithm to support information dissemination and retrieval in Internet-scale Publish/Subscribe system. They also compare 6 different caching strategies in the aspects of user experience, network overhead, etc. Sourlas et al. [16] propose a Publish/Subscribe system supporting mobile users through caching. It enables the mobile users to retrieve missed information generated during their roam. These caching algorithms have a similar approach that they cache the messages in the border brokers (the brokers directly connected by clients) in Publish/Subscribe network. They make use of the existing protocols in traditional Publish/Subscribe network to provide easy access to historical messages while only a small amount of additional protocols is needed. However, in many scenarios, there is a large amount of inner brokers with storage capacity. The above caching algorithms will limit the total storage capacity of the Publish/Subscribe network as they only exploit the storage resources from the border brokers. In the caching algorithm presented in this paper, each broker who has storage capacity will be a possible candidate as a caching point, so the total persistence capacity can be exploited effectively.

Sourlas et al. [17] propose a new storage placement and replication algorithm, which differentiates classes of content, and minimize the clients' response latency in topic-based Publish/Subscribe networks. Sourlas et al. [18] propose a new message storage and replica allocation algorithm. The algorithm stores the messages and allocates the replicas based on the popularity of the message content. It can effectively reduce the user response delay and network overhead while processing user requests. The caching algorithms assume the system has a priori knowledge of location, interest, and frequency of user requests. Then, they place the message replicas close to the users logically or physically, which is similar to the caching algorithms in CDN. However, in most scenarios, it is hard to know the preference of the users who join and leave the system dynamically. In this paper, we will focus on the data persistence without knowing the preference of new users.

3 Caching Algorithm

In this section, we will propose a caching algorithm in the basic Publish/Subscribe network to support message persistence. Our goals are to (1) maximize the storage capacity of the Publish/Subscribe network; (2) reduce the overhead while retrieving the historical messages; (3) make the protocols as simple as possible.

We assume that the publishers are static, which means they will always be online and their position is unchanged; the subscribers are dynamic, which means they can join the system at any time, any place. This assumption is satisfied in many practical scenarios, such as monitoring system on stock market, monitoring system of device data in railway. Meanwhile, we assume that the brokers not only

provide the routing function, but also provide a certain amount of storage capacity, which can be easily implemented by attaching storage devices to the brokers.

In our caching algorithm, we will make use of the storage capacity of the message delivery path. While the message is routing in the delivery path, a caching point (a broker) is calculated using the content of message and a predefined hash function. Then, the message will be stored in that caching point. Likewise, while a user is requesting for the historical messages, the expected caching points are calculated using the request filter and the same predefined hash function.

3.1 Caching Messages

We assume that the message content is composed of a single continuous attribute. We use a hash function to calculate the caching point in the delivery path for each message. The domain of the hash function is the domain of the message content, while the codomain is a set of possible broker hops along the delivery path.

We also assume that if a user requests for the historical messages, it usually sets a continuous range as the content filter. In the design of hash function, we should try to map continuous attribute values to less caching points, as less expected caching points mean less overhead while searching historical messages.

Assumed that a publisher is P_1 and its topic is t_1 (as the message has single attribute, we use the attribute to represent its topic), the attribute t_1 follows the uniform distribution among (A, B) ; the maximum length of delivery path in the broker network is MAX_HOPS. Then, we can design the hash function for caching points as follows:

$$h(e) = \begin{cases} \left\lceil \frac{(t_1.value - A)}{B - A} * (MAX_HOPS - 1) \right\rceil + 1, & A < t_1.value < B \\ MAX_HOPS, & \text{other} \end{cases} \quad (1)$$

When a publisher publishes a message e , e will know it should be stored after $h(e)$ hops in the delivery path. While e is routing along the delivery path, it will maintain a hop counter and be stored in the broker which is at the $h(e)$ hop along the delivery path. Note that when the actual length of delivery path l is less than $h(e)$, our solution is to route backwards $h(e)\%l$ hops from the endpoint of delivery path and store the message there.

For example, as shown in Fig. 2, we assume the topic is “temperature” and it follows the uniform distribution among $(0,100)$. The maximum length of delivery paths is 4. Then, the caching point hash function is as follows:

$$h(e) = \begin{cases} \lceil temp.value/33 \rceil + 1, & 0 < temp.value < 100 \\ 4, & \text{other} \end{cases} \quad (2)$$

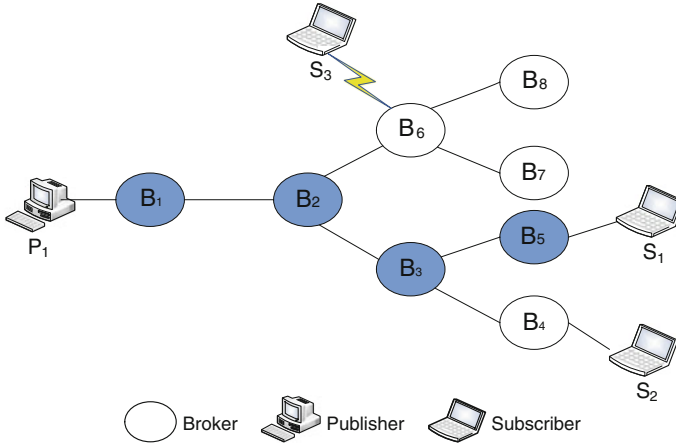


Fig. 2 Publish/Subscribe network supporting data persistence

If P_1 publishes a message e_1 (temperature = 80), the message e_1 will be stored in the 3rd hop ($h(80) = 3$) in the delivery path. If P_1 publishes message e_1 (temperature = -1), the message e_1 will be stored in the 4th hop ($h(-1) = 4$). If the actual length of delivery path of message e_2 is $l = 3$ (hops), the message e_2 will route backwards 1 hop ($h(e_2) \% l$) after reaching the endpoint of its delivery path, and stores the message there.

3.2 Hash Collision

When the storage capacity of a broker run out, the future messages mapped to this broker cannot be stored successfully (it is called hash collision). Our solution is to rehash using linear exploration. When a message e arrives at its expected caching broker b_1 , but the broker b_1 is run out of storage, (1) message e will be stored in the cyclic next hop b_2 ; (2) a tag will be added at broker b_1 , telling that we should continue to do linear exploration while retrieving messages mapped to b_1 ; (3) if both the storage of brokers b_1 and b_2 run out, message e will continue to do linear exploration until it arrives at a broker with available storage.

3.3 Drop Policy

Since the storage resource of broker is limited, hash collisions will occur when brokers run out of storage finally. The hash collision will result in significant increase in overhead while retrieving historical messages. So, we should provide drop policy to clear the storage of brokers.

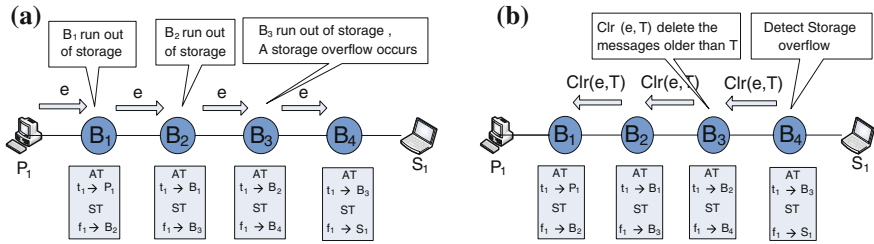


Fig. 3 Clear procedure of delivery path. **a** Storage overflow. **b** Clear delivery path

Once many brokers run out of storage in the delivery path, a great many explorations will occur. This will cause terrible performance decline while retrieving historical messages. We can set the maximum exploration hops to K , which can be adjusted according to the scale of network. If a message explores more than K hops to find its caching broker, it is treated as a storage overflow of the delivery path.

When a storage overflow is detected by a message e , an overflow tag will be added in the message e . Finally, the endpoint broker of the delivery path will detect the storage overflow by checking the overflow tag and a $Clr(e, T)$ message will be generated to start a clear mechanism. The message $Clr(e, T)$ will be routed backwards along the delivery path of message e .

The parameter T means a time threshold. While $Clr(e, T)$ is routing along the delivery path, each broker in the path will delete its historical messages whose timestamps are older than T . When the $Clr(e, T)$ arrives at the beginning of the delivery path, the message e will be delivered and stored as before.

For example, as shown in Fig. 3a, we set the maximum exploration hops $K = 2$. P_1 publishes message e , which is mapped to B_1 . When message e arrives at B_1 , it continues to explore new caching broker as B_1 run out of storage. When both B_2 and B_3 run out of storage, the exploration hops will exceed $K = 2$ and the message e detects a storage overflow of the delivery path.

As shown in Fig. 3b, B_4 finds that there is a storage overflow. A message $Clr(e, T)$ is routed backwards along the routing path of e . Each broker will delete its historical messages older than T . Finally, when $Clr(e, T)$ arrives at the beginning of the delivery path, message e will continue to be delivered and stored as before.

3.4 Request and Response Mechanism

In this session, we will add the request and response mechanism to support the retrieval of historical messages.

If a new subscriber wants to retrieve historical messages, it will send out a message $Request(f)$, in which f is the content filter for historical messages. $Request(f)$ will be routed to the dedicated beginning point of the delivery path

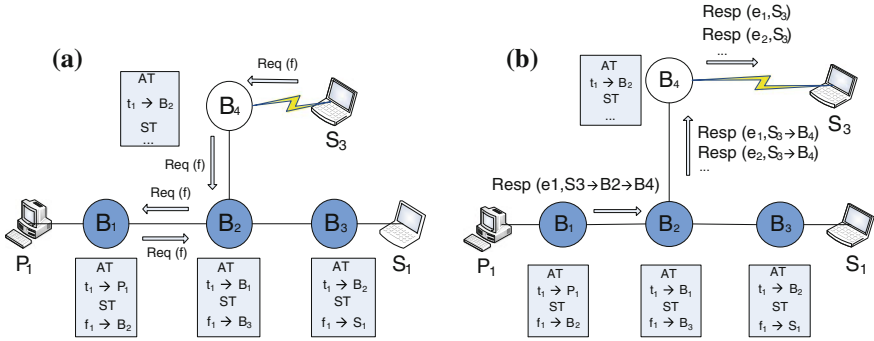


Fig. 4 Request and response mechanism. a Request. b Response

according to the topic related to f and the advertisement table (AT). At the beginning point, the set of expected caching points can be calculated using the predefined hash function. Then, Request(f) will be routed in the same way as messages along the delivery path, using filter f to match with the filters in the subscription table (ST).

While Request(f) is routing along the path, it will look up the storage in the expected caching points. If a historical message e matching the filter f is found, a Response(e) message will be created and routed backwards along the Request(f)’s routing path. Finally, the Response(e) will be routed to the subscriber.

For example, as shown in Fig. 4a, when a new subscriber S_3 wants to retrieve the historical messages matching its interest, it sends out a Request(f), in which f means the filter of historical messages. Here, the filter f belongs to topic t_1 ; filter f and f_1 intersect, which means: $\{e | e \text{ matches } f\} \cap \{e | e \text{ matches } f_1\} \neq \Phi$.

We assume the set of expected caching points is $\{1\}$ and B_1 has a tag for f_1 , telling that we should continue exploration while retrieving messages matching f_1 . First, Request(f) is routed to B_1 according to the advertisement table (AT). Then, Request(f) is routed according to the subscription table (ST). When Request(f) arrives at the B_1 , it will look up its storage for historical messages matching f . Meanwhile, the B_1 has a exploration tag for filter f , so Request(f) continues to be routed to B_2 and look up for historical messages there. As B_2 does not have any exploration tag for filter f , the Request(f) will not be routed any further and the request procedure is done.

As shown in Fig. 4b, if a historical message e_1 matching f is found in B_1 , a Response(e_1) message is generated. As the recorded routing path of Request(f) is $S_3 \rightarrow B_4 \rightarrow B_2 \rightarrow B_1$, we set the routing path of Response(e_1) to $B_1 \rightarrow B_2 \rightarrow B_4 \rightarrow S_3$ and route the Response(e_1) to S_3 . If a historical message e_2 matching f is found in B_2 , a Response(e_2) message is generated. While the recorded routing path of Request(f) is $S_3 \rightarrow B_4 \rightarrow B_2 \rightarrow B_1 \rightarrow B_2$, we can set the routing path of Response(e_1) to $B_2 \rightarrow B_4 \rightarrow S_3$, so as to avoid unnecessary loops.

4 Experiments and Evaluation

In this chapter, we simulate the Publish/Subscribe system and implement the proposed caching algorithm. We use three computers connected through Ethernet, each equipped with 2.6-GHz CPU and 4G RAM. Each computer runs several broker processes. In the experiment of [16], the overlay network is organized as a balanced binary tree, which can avoid the bias of overhead between brokers. So, we will also adopt this topology of broker network, in which publisher is connected to the root of tree, while the subscribers randomly appear in the leaves of the tree.

We assume that the message contains a single continuous attribute which follows the uniform distribution among a specified interval. We will also use a continuous interval as the message filter and use the proportion of the interval's length to the attribute's total length as the coverage rate of filter, marked as R . The number of brokers is marked as $N(\text{brokers})$, the cache size of broker is marked as $K(\text{messages})$, and the publish rate of publisher is marked as $S(\text{messages/second})$.

The performance metrics we observe are as follows:

- (1) Lifetime of Message. The interval from the time when the message is published to the time when the message disappears from the network, which represents the persistence capacity of the network.
- (2) Searched Cache Slots. The total volume of the searched cache slots during request procedure, which represents the overhead in retrieving historical messages.
- (3) Response Delay of Historical Message. The time delay of the response procedure, which represents the retrieving speed of historical messages.

We compare our caching algorithm with other two algorithms. Our caching algorithm is named as path caching using hash mapping (PCHM), and the caching point hash function is shown in formula (1). Another classic caching algorithm is end point caching (EPC), which stores the message in the end point of the delivery path [16]. The third caching algorithm is a variant of PCHM, and the only difference is that the hash function is changed to $h(e) = t.\text{value} \% \text{MAX_HOPS}$. The MAX_HOPS means the maximum hops of the current delivery path. This algorithm is named as path caching using hash mapping 2 (PCHM2).

Figure 5 depicts the average lifetime of message, which reflects the persistence capacity of the system. In Fig. 5a, as the publish rate increases in the network, the persistence capacity of the system declines in logarithmic level. PCHM and PCHM2 have a better performance in persistency capacity. In Fig. 5b, as the cache size of broker increases, the persistence capacity of the system increases in linear level. The persistency capacity of system using PCHM and PCHM2 grows faster than that of the system using EPC. In Fig. 5c, as the number of brokers increases, the persistence capacity of the system using EPC does not change, while the persistence capacity of the system using PCHM and PCHM2 grows in logarithmic level. In general, PCHM and PCHM2 will perform better than EPC in persistency capacity.

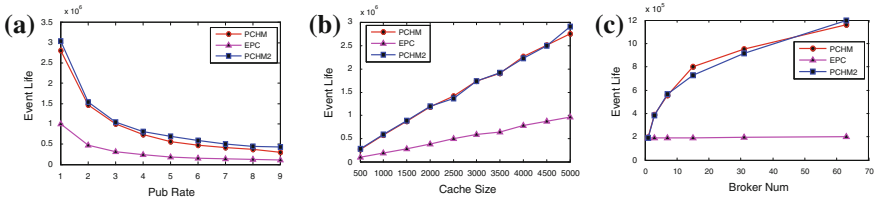


Fig. 5 Lifetime of messages in the network

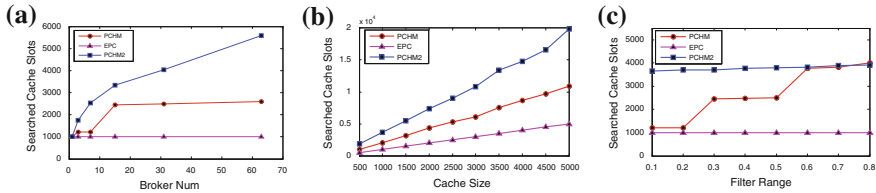
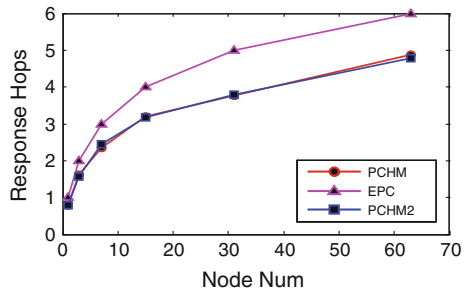


Fig. 6 Searched cache slots of request procedure

Figure 6 depicts the changes in searched cache slots, which represents the overhead in retrieving historical messages. Here, we assume that the coverage rate of filter for historical messages is 0.3 ($R = 0.3$). In Fig. 6a, as the number of brokers increases, the searching overhead of EPC remains unchanged, while the searching overhead of PCHM2 grows in logarithmic level. The searching overhead of PCHM is between EPC and PCHM. Given that the filter coverage rate remains unchanged, as the number of brokers increases, the searching overhead of PCHM almost remains unchanged after some point. In Fig. 6b, we also assume that the coverage rate of filter for historical messages is 0.3 ($R = 0.3$). As the cache size increases, the overhead of searching historical messages grows in linear level, while the speed of growth of PCHM increases, and the searching overhead of EPC and PCHM2 almost remains unchanged, while the searching overhead of PCHM is between EPC and PCHM2. In general, PCHM and EPC perform better than PCHM2 in the overhead of searching historical messages.

Figure 7 depicts the response delay of historical messages, which is an important metric in representing the quality of service from the user’s perspective.

Fig. 7 Response delay of historical data



In Fig. 7, as the number of brokers increases, the response delay grows in logarithmic level. The algorithms PCHM and PCHM2 perform better than EPC in the response delay of historical messages.

5 Conclusion

This paper proposes a distributed message caching protocol in Publish/Subscribe network and provides request and response mechanisms for retrieving historical messages. So, we can enable the users to subscribe to the historical messages in Publish/Subscribe system. Comparative experiments with the other two caching algorithms are conducted, and the result shows that the proposed caching algorithm provides the system with good persistency capacity, low overhead of history retrieval, low user response delay, and good scalability.

The Publish/Subscribe network with message persistency can be applied to many scenarios. For example, with the popularity of mobile computing, many clients join the system dynamically. The Publish/Subscribe network with message persistency can be extended to support client mobility, as the mobile client can retrieve the messages missed during roam. Another example is that the Publish/Subscribe network with message persistency can be extended to provide reliable message delivery, as we can retrieve the message while the message missing is detected.

References

1. Eugster PT, Felber PA, Guerraoui R, Kermarrec AM (2003) The many faces of publish/subscribe. *ACM Comput Surv* 35:114–131
2. Yuan HL (2006) Research on key technologies for supporting content-based publish/subscribe, National University of Defense Technology, Changsha, p 12
3. IBM TJ (2001) Watson research center. Gryphon: publish/subscribe over public networks. <http://researchweb.watson.ibm.com/gryphon/Gryphon>
4. Carzaniga A (1998) Architectures for an event notification service Scalable to wide-area networks. PhD thesis, Politecnico di Milano, Milan, Italy
5. Carzaniga A, Rosenblum DS, Wolf AL (2001) Design and evaluation of a wide-area event notification service. *ACM Trans Comput Syst* 19(3):332–383
6. Cugola G, Di Nitto E, Fuggetta A (2001) The JEDI event-based infrastructure and its application to the development of the OPSS WFMS. *IEEE Trans Softw Eng* 27(9):827–850
7. Fiege L, Mühl G (200) Rebeca event-based electronic commerce architecture. <http://event-based.org/rebeca>
8. Pietzuch PR (2004) Hermes: a scalable event-based middleware. PhD thesis, University of Cambridge, Cambridge, United Kingdom
9. Segall W, Arnold D (1997) Elvin has left the building: a publish/subscribe notification service with quenching. In: Proceedings of the 1997 Australian UNIX Users Group, Brisbane, Australia, pp 243–255. <http://elvin.dstc.edu.au/doc/papers/auug97/AUUG97.html>

10. Baldoni R, Contenti M, Piergiovanni ST, Virgillito A (2003) Modeling publish/subscribe communication systems: towards a formal approach. In: Proceedings of the eighth international workshop on object-oriented realtime dependable systems, 2003. (WORDS 2003), Issue 15–17, pp 304–311
11. Pierre G, van Steen M (2006) Globule: a collaborative content delivery network. *IEEE Commun* 44(8):127–133
12. Li G, Cheung A, Hou S, Hu S, Muthusamy V, Sherafat R, Wun A, Jacobsen H, Manovski S (2007) Historic data access in publish/subscribe. In: Proceedings of the 2007 inaugural international conference on distributed event-based systems (DEBS 2007), Toronto, Canada, pp 80–84
13. Singh J, Eyers DM, Bacon J (2008) Controlling historical information dissemination in publish/subscribe. In: Proceedings of the 2008 workshop on middleware security (MidSec2008), Leuven, Belgium, pp 34–39
14. Sourlas V, Paschos GS, Flegkas P, Tassiulas L (2009) Caching in content-based publish/subscribe systems, to appear in *IEEE Globecom 2009 next-generation networking and internet symposium*
15. Diallo M, Fdida S, Sourlas V, Flegkas P, Tassiulas L (2011) Leveraging caching for Internet-scale content-based publish/subscribe networks. In: Proceedings of international conference communication 2011 (*IEEE ICC2011*), pp 1–5
16. Sourlas V et al (2010) Mobility support through caching in content-based publish/subscribe networks. In: 10th *IEEE/ACM international conference on cluster, cloud and grid computing*, pp 715–720
17. Sourlas V, Flegkas P, Paschos GS, Katsaros D, Tassiulas L (2010) Storing and replication in topic-based publish/subscribe networks. In: Proceedings of *IEEE Globecom*, Miami, USA
18. Sourlas V, Flegkas P, Paschos GS, Katsaros D, Tassiulas L (2011) Storage planning and replica assignment in content-centric publish/subscribe networks. *Comput Netw* 55(18):29

Vision-Based Environmental Perception and Navigation of Micro-Intelligent Vehicles

Ming Yang, Zhengchen Lu, Lindong Guo, Bing Wang
and Chunxiang Wang

Abstract The adjustment of actual environmental traffic flow experiments is complicated and time-consuming. In this paper, a method based on micro-intelligent vehicles (micro-IV) is proposed to overcome these unfavorable factors. Vision-based environmental perception employed should be qualified for real-time and robust characteristics. An active vision approach based on visual selective attention is carried out to search regions of interest more efficiently for traffic light recognition. Corner detection based on histogram is applied for real-time location in autonomous parking. A method based on hierarchical topology maps is proposed to realize the navigation without GPS equipment. Experimental results show that the perception and navigation approaches work efficiently and effectively and micro-IV is suitable for traffic flow experiments.

Keywords Micro-intelligent vehicles · Traffic flow simulation · Traffic light recognition · Autonomous parking

1 Introduction

In contemporary times, traffic congestion has increasingly become one of the worldwide pressing issues. Scientific management and control of traffic flow is considered as one of the most cost-effective strategies to implement in urban traffic environments. The initial issue in the traffic flow management and control is the

M. Yang · Z. Lu (✉) · L. Guo · B. Wang

Department of Automation, Shanghai Jiao Tong University, and Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai 200240, China

e-mail: general_zclu@hotmail.com

C. Wang

Research Institute of Robotics, Shanghai Jiao Tong University, Shanghai 200240, China

e-mail: wangcx@sjtu.edu.cn

modeling and simulation of traffic flow. Most of relative research works are based on software simulation [1, 2]. However, there is still considerable controversy in several facts about whether software simulation matches real traffic flow enough for its shortages including rare dynamic traffic simulation based on path planning, insufficient analysis of intelligent traffic system, and cooperative vehicle infrastructure system equipped with wireless network communication.

For the limitation of software simulation, the importance of experiments in real traffic environments is being concerned in traffic flow research [3]. However, the adjustment of real traffic flow experiments is complicated and time-consuming [4]. Also, there hardly ever exists chance for multiple intelligent vehicles involved in. As a consequence, a method based on multiple micro-intelligent vehicles (micro-IV) is proposed for traffic flow simulation which can be freed from weather, law, and other factors which trouble real traffic flow experiments and autonomous navigation driving experiments in real urban traffic environments.

Micro-IV method can bridge over these difficulties by creating a microproportional traffic flow environment that is designed with different kinds of complex traffic scenes, in which T-junction, crossroad, and various styles of lanes are involved. In order to realize the traffic flow simulation, a certain quantity of micro-intelligent vehicles is required to drive in the microtraffic environments autonomously and stably without the help of GPS for its microscale. Also taking the micro-IV expense into consideration, it is unrealistic to arm such number of micro-IV with expensive high-precision sensor like laser scanner, which is widely used for obstacle detection in real intelligent vehicles. However, these micro-intelligent vehicles should be capable of different driving tasks as the same that real vehicles possess including driving in lane with collisions avoidance and obey traffic rules, driving through intersection with traffic light recognition, autonomous parking with parking area detection, and autonomous navigation with topology map. The above tasks mostly rely on vision information obtained by two cameras placed on micro-IV. So, vision-based environmental perception and navigation method employed in this situation play a fundamental role in micro-IV system.

Both real time and accuracy are essential characteristics and requirements of micro-IV environmental perception method. Although some researches on traffic light recognition have already been carried out in different occasions [5–15], it is hardly to search out a method that satisfies the case of micro-IV. Reference [9] proposed an approach to acquire traffic light area that illumination parameters needed to be estimated which is unsuitable for micro-IV. Also, the algorithm in [10] is applied for a static situation because the camera is provided with a fixed position. The computation amount of the approaches in [12] is too large to meet the demand of real-time detection. In fact, these approaches are all developed in passive vision mode which means they are organized in bottom-up attention manner via a series of pretreatment and image processing and image segmentation to fulfill the detection. Generally, bottom-up attention mode can reach good results for feature extraction in local significant region of image, while lack of consideration of the global image statistics information makes bottom-up attention fail to

turn attention to region of interests (ROI) rapidly, which leads to the fact that real time and accuracy cannot be reconciled well for bottom-up attention methods.

However, it is obvious that humans can instantly identify the most relevant ROI from global vision based on different tasks through active vision mode (i.e., top-down attention mode) and only need to analyze certain parts of the image which strongly reduces the amount of computation and improves the accuracy [16–20]. If the computable and swift model of active vision can be available, then image feature extraction will be much improved. An approximate computable top-down attention mode called visual selective attention (VSA) is proposed in [16, 18] which determines ROI through weighted average saliency maps of different features including intensity feature, RGB feature, and Gabor filtering feature. In this paper, a pretreatment is applied to covert RGB to HSI. HSI color mode is chosen as the features for traffic light recognition instead of RGB color mode.

Parking area detection is carried out based on contour and corner detection instead of VSA. Gabor filtering feature mentioned in VSA is a kind of human-like algorithm to search direction and texture information for active vision [17, 18], while Gabor filtering is rather computationally intensive which are not suitable for micro-IV [18]. Most parking area detection methods proposed are based on distortion correction and orthographic projection [21]. But distortion correction and orthographic projection transform will occupy certain period of time, so that it is not feasible for micro-IV. Also, orthographic projection transform will be useless if the object region is far away from the viewpoint [21]. It requires the ground to be smooth and horizontal which is too idealized in reality. In this paper, a robust and real-time method will be used to detect parking area.

The following paper is organized as follows: Sect. 2 will give detail description of recognition method of traffic lights. Parking area detection method will be discussed in Sect. 3. Navigation method for micro-IV will be introduced in Sect. 4. The experimental results and conclusions are given in Sects. 5 and 6.

2 Traffic Light Recognition

Traffic light recognition approach employed in micro-IV is divided into a series of procedures. First, a preprocessing of image is applied to form the image suitable for VSA which includes an image resize process and a color space transform from RGB to HSI. Then, VSA is carried out to generate various feature saliency maps which are used to get ROI through weighted average afterward. Next, position of ROI for origin image will be calculated proportionally from that of resized image. In the end, a bottom-up detection method will be carried out to achieve the ultimate recognition results.

Color information is the most important information for traffic light recognition. RGB color components are not mutually independent which means it is difficult to distinguish different hues in RGB color space, while, as is widely accepted, HSI color space represents color with hue, saturation, and intensity, which is quite

convenient for color detection by defining HSI threshold for certain color [5]. So, color space transform from RGB to HSI cannot be omitted.

The hue threshold of traffic light can be defined as

$$\text{Hue}_{\text{color}} = \{p(x, y) | (H_{\text{color1}} < P_{H_{\text{color}}} < H_{\text{color2}})\} \quad (1)$$

where color can be replaced with red, green, and yellow [5]. H_{color1} is the min threshold of hue, H_{color2} is the max threshold of hue, and $P_{H_{\text{color}}}$ is the hue of every pixel. To emphasis on traffic light color from others, a hue band-pass filtering based on normal density function is applied. The function that is

$$P'_{H_{\text{color}}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(P_{H_{\text{color}}}-\mu)^2}{2\sigma^2}} \quad (2)$$

where μ represents the center position corresponding to the peak, and σ represents the steepness of the function curve. For instance, if the red color wanted to be stressed out, μ should be appointed to the center threshold of red hue. The bigger σ is, the more obvious other adjacent color will be. After that, the gray-scale image of hue (H_{gs}) is generated from $P'_{H_{\text{color}}}$ image, of which the pixels are normalized to ensure that the values meet the domain.

Then, a spectral residual approach [16, 18] is carried out to produce a hue saliency map (HSM) from H_{gs} , which is qualified with the characteristic of real time. In this approach, image information is divided into innovation and redundancy. Once the redundancy part is removed, the innovation part will be preserved and the search region can be reduced. The overall information of image is described with log-spectrum, and the redundancy information is extracted by convolution between mean filter template and log-spectrum. After that, the innovation part can be obtained and a HSM feature can be produced by inverse fast Fourier transform.

In this procedure, it is feasible to get saturation saliency map (SSM) and intensity saliency map (ISM) of the image. While there is a difference between the effect of saturation and intensity, SSM produces a relatively concentrated attention result on traffic light. In contrast, ISM produces a distraction result on traffic light.

As a result of these facts, except ISM, HSM, and SSM are selected to integrate into a chief saliency map (CSM) through weighted average fusion method [17, 18], as shown in Fig. 1a. The brightest region pointed out the ROI of traffic light.

$$P_{\text{CSM}} = \omega_1 * P_{\text{HSM}} + \omega_2 * P_{\text{SSM}} \quad (3)$$

After the position of ROI is confirmed, a series of bottom-up recognition method based on shaped feature filtering are carried out to achieve recognition result [5], as shown in Fig. 1b. Some other experimental results including microtraffic environment and actual traffic environment are displayed in Figs. 2, 3, and 4.

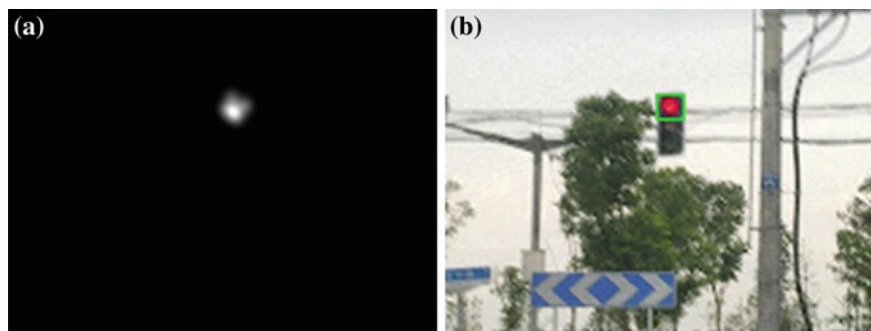


Fig. 1 CSM image (a) is the weighted average image of HSM and SSM. Then, bottom-up recognition result is marked with *green rectangle* in the origin image (b)

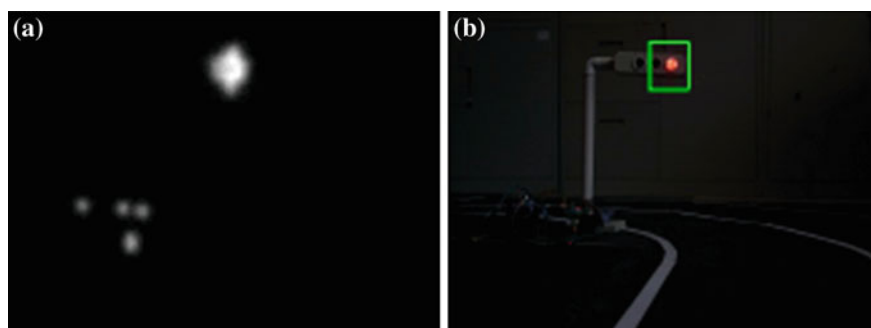


Fig. 2 CSM (a) and recognition results of *red light* (b)

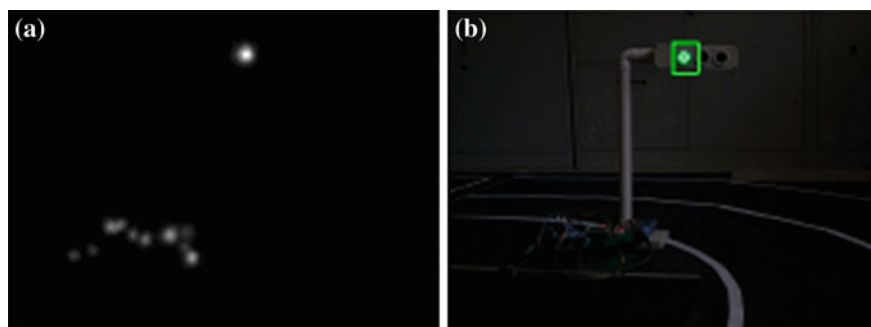


Fig. 3 CSM (a) and recognition results of *green light* (b)

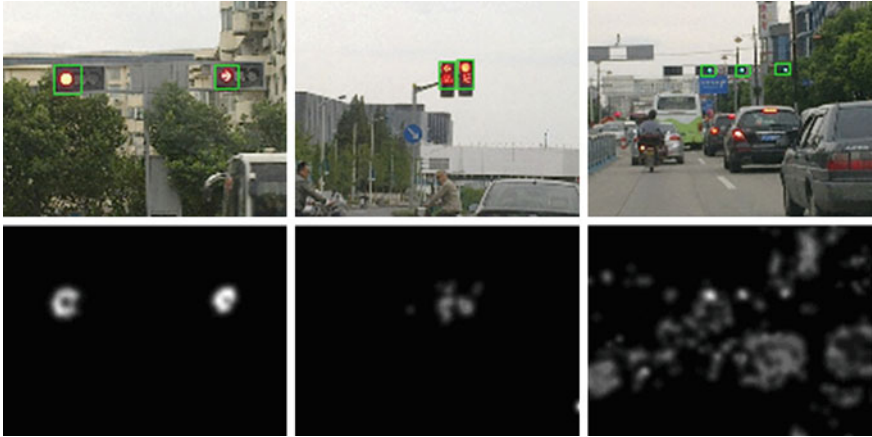


Fig. 4 Some other traffic light recognition results and corresponding CSM

3 Parking Area Detection

Because parking area is consisted of various parallelograms, saliency map based on simple features (e.g., color, intensity) is not distinguished enough for parking area detection. Direction features based on Gabor filtering [17, 18] also cannot pick out the parking area quickly and correctly enough. In order to overcome the difficulties, a pretreatment based on contour detection is applied to realize an extraction of ROI. When a suitable contour comes into view, parking corner detection will be carried out to locate the parking position.

Firstly, Otsu method is applied to find a suitable threshold for generation of binary image. Then, hole contour candidates are checked out with regard to area and perimeter threshold, as shown in Fig. 5a. Dimension ratio (DR) is also taken into consideration to remove interference [5].

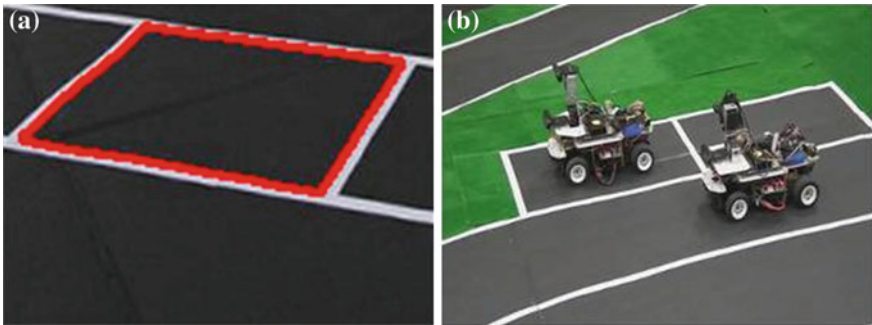


Fig. 5 Contour detection (a) is carried out for pretreatment. Then, the attitude of camera (b) will change to find parking area corner

One of the significant features of parking area is the parking corner. In most of the time, it is hard for vehicles or micro-IV to achieve panoramic image of parking space. So, it is necessary to detect the parking corner to determine the relative position between vehicles and parking space. In this part, a corner detection method based on histogram is applied.

Firstly, perspective of camera is rotated to form an ortho-photo of road as shown in Fig. 5b. Binary image is established with Otsu method. Then, edge detection based on Canny operator is applied to get the edge image from binary image. Secondly, Radon transform [22] is carried out for edge image within the angles range from $[90^\circ - \theta, 90^\circ + \theta]$ where the main direction exists in order to check the inclination of the image, as shown in Fig. 6. Then, the image is rotated by a certain angle to make the main direction parallel with horizontal axis, as shown in Fig. 7. Thirdly, the distribution of white pixels of binary image is counted both in rows and columns, as shown in Fig. 8. The center of corner can be calculated as followed:

$$x_{\text{corner_center}} = \frac{\sum_{i=1}^{\text{width}} x_i \cdot \delta(x_i)}{\sum_{i=1}^{\text{width}} \delta(x_i)}, \quad y_{\text{corner_center}} = \frac{\sum_{i=1}^{\text{height}} y_i \cdot \delta(y_i)}{\sum_{i=1}^{\text{height}} \delta(y_i)} \quad (4)$$

$$\delta(x_i) = \begin{cases} 1, & g(x_i) > T_x \\ 0, & \text{else} \end{cases}, \quad \delta(y_i) = \begin{cases} 1, & g(y_i) > T_y \\ 0, & \text{else} \end{cases} \quad (5)$$

where $g(x_i)$ represents the number of white pixels in x_i column, T_x is the threshold of $g(x_i)$, and $\delta(x_i)$ has only two status. Similarly, $g(y_i)$ represents the number of white pixels in y_i row, and T_y is the threshold of $g(y_i)$. In this way, the corner center coordinate can be found, as shown in Fig. 9b. Then, micro-IV can locate the relative position between the parking space and itself with the knowledge of corner position and image inclination angle.

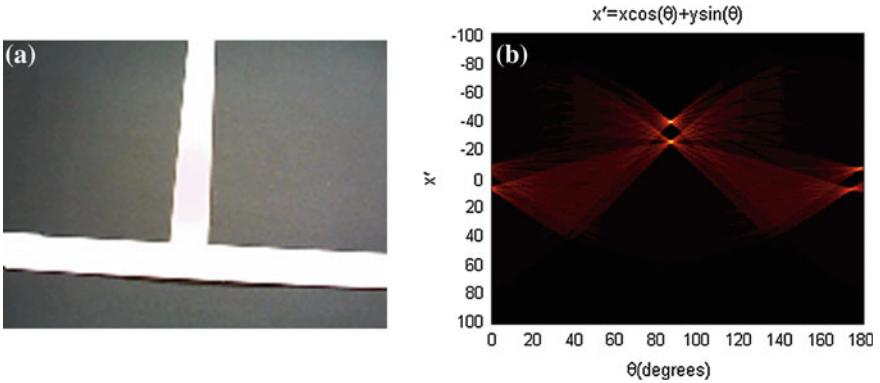


Fig. 6 Origin image of corner (a) may be oblique, and Radan transform (b) is applied to check the inclination angle

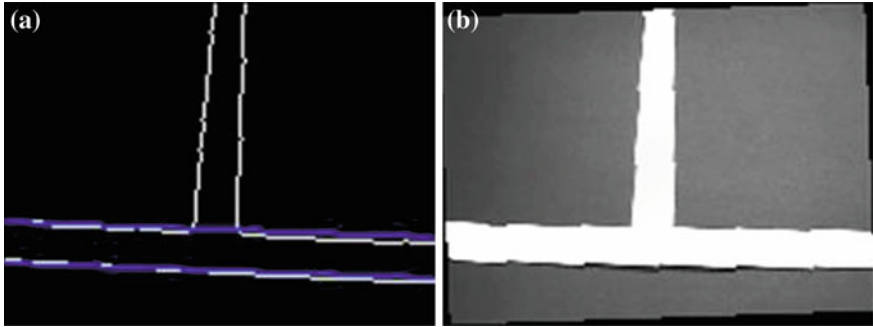


Fig. 7 The main direction of lane can be found (a), and the image is rotated to be parallel to the axis (b)

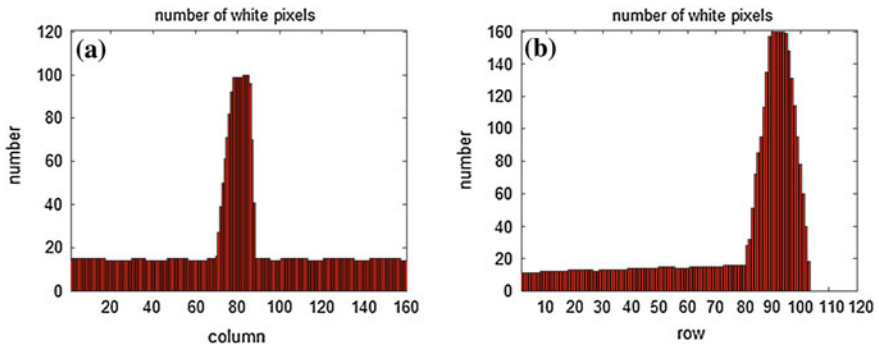


Fig. 8 White pixels of binary image are counted in columns (a) and rows (b)

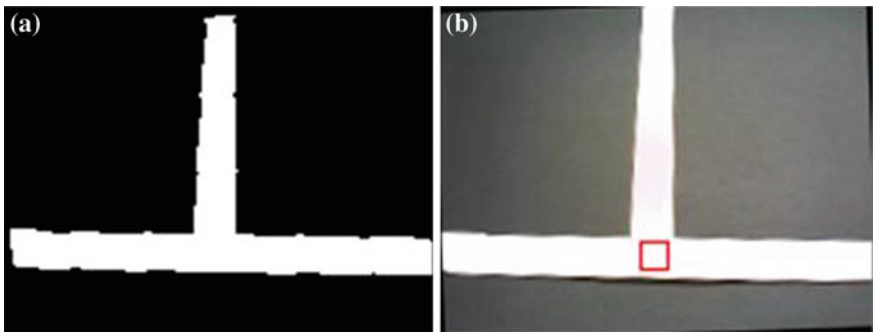


Fig. 9 The rotated image is converted into binary image (a), and the corner can be found with histogram that is marked in rotated image (b)

4 Navigation with Topology Map

Due to lack of global location information, a navigation method based on topology map like early times of human driving is proposed. The topology map is designed in two layers. In first topology map layer, the micro-urban traffic environment, as shown in Fig. 10, is simplified as the classic graph theory. That first topology map layer is described as a tuple $M_1 = (E,P)$, where $E = \{e_1,e_2,\dots,e_n\}$ is a nonempty set of nodes of which e_i represents a single intersection and $P = \{p_1,p_2,\dots,p_n\}$ is another nonempty set of directed edges, of which p_i represents a single lane. There are some special circumstances should be defined to ensure consistency. (1) To end-to-end loop lane, a virtual intersection is defined on the lane. (2) The end of the blind alley is also defined as an intersection which can define U-turn direction in second topology map layer.

Then, an adjacency matrix can be derived to describe M_1 as

$$C_{ij} = \begin{cases} 1, & (e_i, e_j) \in P \\ 0, & \text{else} \end{cases} \tag{6}$$

which is incapable of navigation for micro-IV after passing an intersection because an intersection may connect to multiple lane entrances. So, a detail map for intersection is necessary. A tuple $M_2 = (R,S)$, where $R = (r_1,r_2,\dots,r_n)$ is a nonempty set of nodes of which r_i represents a single lane connected to the intersection, and there exists $R = P$. $S = (s_1,s_2,\dots,s_n)$ is a nonempty set of edges of which s_i represents a single permitted driving action in certain intersection. Then, an adjacency matrix can be applied to describe M_2 as

$$C_{ij} = \begin{cases} 1, & (r_i, r_j) \in S \wedge (r_i, r_j) \in T_{\text{Left}} \\ 2, & (r_i, r_j) \in S \wedge (r_i, r_j) \in T_{\text{Right}} \\ 3, & (r_i, r_j) \in S \wedge (r_i, r_j) \in T_{\text{Straight}} \\ 4, & (r_i, r_j) \in S \wedge (r_i, r_j) \in T_{U\text{-Turn}} \\ 0, & \text{else} \end{cases} \tag{7}$$



Fig. 10 Microtraffic environments

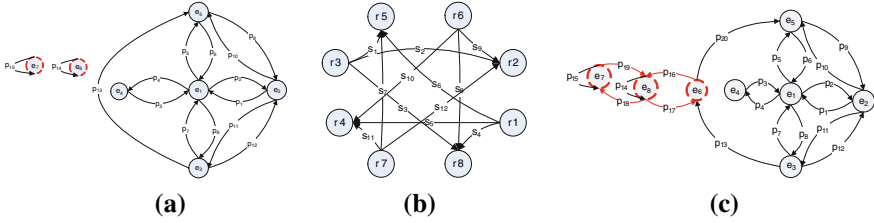


Fig. 11 Basic topology map M_1 is shown as (a), and topology map M_2 is shown as (b). Extended topology map M_1 is shown as (c). Red nodes refer to the virtual intersection. Red edges refer to the lane change actions

where T_i ($i = \text{Left, Right, Straight, U-Turn}$) represents different driving direction in the intersection. M_1 and M_2 are shown as Fig. 11a and b. However, the position still cannot be available when a lane-changing action occurred. The definition of M_1 and M_2 must be extended to fit this situation. Those lanes have the permission of changing lane will increase a virtual intersection except those that already have one. Adjacency matrix of M_2 can be extended as

$$C_{ij} = \begin{cases} 1, (r_i, r_j) \in S \wedge (r_i, r_j) \in T_{\text{Left}} \\ 2, (r_i, r_j) \in S \wedge (r_i, r_j) \in T_{\text{Right}} \\ 3, (r_i, r_j) \in S \wedge (r_i, r_j) \in T_{\text{Straight}} \\ 4, (r_i, r_j) \in S \wedge (r_i, r_j) \in T_{U\text{-Turn}} \\ 5, (r_i, r_j) \in S \wedge (r_i, r_j) \in T_{\text{Left-Change}} \\ 6, (r_i, r_j) \in S \wedge (r_i, r_j) \in T_{\text{Right-Change}} \\ 0, \text{else} \end{cases} \quad (8)$$

where T_i adds left and right lane-changing actions into intersection actions. And extended M_1 is shown as Fig. 11c. With the help of these two topology maps, micro-IV can locate itself within the lane level once the initial position is fixed. In micro-IV system, these two topology maps are stored in form of entity-relationship (E-R) model because the form of adjacency matrix is not a space-saving model. The E-R model of map M_1 is a tuple: lane(P, SE, EE), where P is the lane set mentioned in map M_1 . SE is the set of start node of lane, EE is the set of end node of lane, and $SE = EE = E$. E-R model of map M_2 is a tuple: action(S, E, T, FP, NP). Where $S, E,$ and T is the set mentioned above. FP is the set of former lane before intersection action, NP is the set of next lane after action, and $FP = NP = P$.

5 Experiment Results

The given methods have been implemented in our micro-IV experimental platform that is equipped with two web cameras, as shown in Fig. 12. The lower one is responsible for traffic light recognition. The higher one, with two degrees of

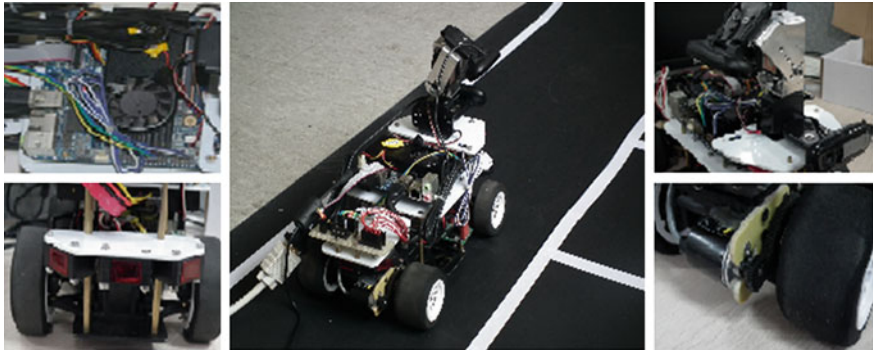


Fig. 12 Micro-IV experimental platform

freedom, is in charge of lane-keep driving and parking area detection. Micro-IV is also installed with a single-board PC for image processing, of which the frequency is 1.6 GHz and processors are AMD G-T56N. The style of main memory is 2G. Also, an optical encoder is employed for speed feedback and an array of infrared distance sensors is deployed for obstacle detection.

In traffic light recognition, the parameters of red hue band-pass filtering are set as $\mu = 0.99$ and $\sigma = 0.01$. Parameters of saturation band-pass filtering are configured as $\mu = 0.95$ and $\sigma = 0.02$. Weighted average fusion is applied with $\omega_1 = \omega_2 = 0.5$. In parking area detection, the contour is selected with area ranging from 5,000 to 5,600. Radon transform parameter $\theta = 10^\circ$. The average computational time is shown in Table 1. The recognition statistics as shown in Table 2 include traffic light recognition recall and parking area detection recall (speed = 0.5 m/s). Some captured images of experiments are shown in Fig. 13. Video of micro-IV can be found at:http://v.youku.com/v_show/id_XNDEzMjAwNTk2.html.

Table 1 Average computational time

	Time (ms)
Traffic light image pretreatment	9.2
Traffic light CSM generation	41.2
Traffic light bottom-up recognition	12.4
Parking contour detection	4.8
Parking corner detection	25.2

Table 2 Recognition statistics

	Recall (%)
Traffic light recognition	947
Parking area detection	87

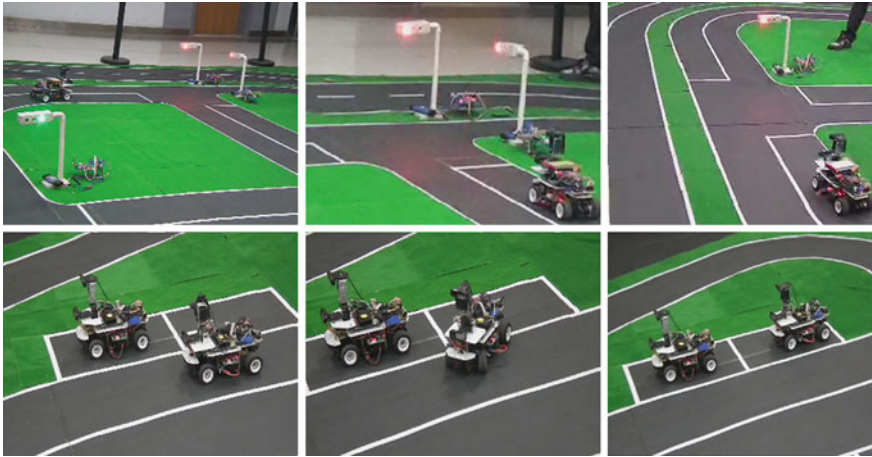


Fig. 13 Some captured images of experiments

6 Conclusions

In this paper, a hardware traffic flow simulation approach based on micro-IV is proposed. VSA method based on HSI features is applied for traffic light recognition to select ROI efficiently and achieve a real time and robust result. Parking area detection based on contour detection and corner detection enables micro-IV to aware parking area in a remote distance and locates the parking space fast and accuracy. Micro-IV is also capable of navigation with the help of topology maps. As a consequence, micro-IV approach is suitable for hardware traffic flow simulation. In future, VSA method will be further researched in the field of traffic sign detection and pedestrian detection. Also, an efficient VSA description of shape and orientation or other complex features will be discussed for micro-IV.

Acknowledgments This work was supported by the Major Research Plan of National Natural Science Foundation (91120018/91120002), the General Program of National Natural Science Foundation of China (61174178/51178268), and National High-tech R&D 863 Program (2011AA040901).

References

1. Yang M, Wan N, Wang B, Wang C, Xie J (2011) CyberTORCS: an intelligent vehicles simulation platform for cooperative driving. *Int J Comput Intell Syst* 4(3):378–385
2. Sun F, Sun Z, Li H (2005) Stable adaptive controller design of robotic manipulators via neuro-fuzzy dynamic inversion. *J Rob Syst* 22:809–819
3. Boxill SA, Yu L (2000) An evaluation of traffic simulation models for supporting ITS development. SWUTC//00/167602-1

4. Sugiyama Y, Fukui M, Kikuchi M, Hasebe K, Nakayama A et al (2008) Traffic jams without bottlenecks-experimental evidence for the physical mechanism of the formation of a jam. *New J Phys* 10:033001
5. Wang C, Tao J, Yang M, Wang B (2011) Robust and real-time traffic lights recognition in complex urban environments. *Int J Comput Intell Syst* 4(6):1383–1390
6. Gong J, Jiang Y, Xiong G (2010) The recognition and tracking of traffic lights based on color segmentation and CAMSHIFT for intelligent vehicles. *IEEE intelligent vehicles symposium, University of California, San Diego, CA, USA*, 431–435
7. Omachi M, Omachi S (2010) Detection of traffic light using structural information. In: *The IEEE international conference on signal processing (ICSP)*, 809–812
8. Lindner F, Kressel U, Kaelberer S (2004) Robust recognition of traffic signals. *IEEE intelligent vehicles symposium, Piscataway, NJ, USA*, 49–53
9. Chung Y, Wang J, Chen S (2002) A vision-based traffic light detection system at intersections. *J Nat Taiwan Normal Univ Math Sci Technol* 47:67–86
10. Yung N, Lai A (2001) An effective video analysis method for detecting red light runners. *IEEE Trans Veh Technol* 50:1074–1084
11. Kim YK, Kim KW, Yang X (2007) Real time traffic light recognition system for color vision deficiencies. In: *IEEE international conference on mechatronics and automation*, 76–81
12. Shen Y, Ozguner U, Redmill K (2009) A robust video based traffic light detection algorithm for intelligent vehicles. *IEEE intelligent vehicles symposium, Piscataway, NJ, USA*, 521–526
13. Charette RD, Nashashibi F (2009) Traffic light recognition using image processing compared to learning processes. In: *The 2009 IEEE/RSJ international conference on intelligent robots and systems, St. Louis USA*, 333–338
14. Charette RD, Nashashibi F (2009) Real time visual traffic lights recognition based on spot light detection and adaptive traffic lights templates. *IEEE intelligent vehicles symposium, Piscataway, NJ, USA*, 358–363
15. Lu K, Wang C, Chen S (2008) Traffic light recognition. *J Chin Inst Eng* 31(6):1069–1075
16. Hou X, Zhang L (2007) Saliency detection: a spectral residual approach. In: *IEEE computer society conference on computer vision and pattern recognition, IEEE computer society, Los Alamitos, CA, USA*, 1–8
17. Itti L, Koch C (2001) Feature combination strategies for saliency-based visual attention systems. *J Electron Imaging* 10(1):161–169
18. Zhang Q, Gu G, Xiao H (2009) Computational mode of visual selective attention. *Robot* 31(6):574–580
19. Navalpakkam V, Itti L (2006) An integrated model of top-down and bottom-up attention for optimizing detection speed. In: *IEEE computer society conference on computer vision and pattern recognition, IEEE computer society, Los Alamitos, CA, USA*, 2049–2056
20. Won WJ, Ban SW, Lee M (2005) Real time implementation of a selective attention model for the intelligent robot with autonomous mental development. In: *IEEE international symposium on industrial electronics (ISIE), IEEE industrial electronics society, Piscataway, NJ, USA*, 1309–1314, June 2005
21. Jung HG, Kim DS, Yoon PJ (2006) Parking slot markings recognition for automatic parking assist system. In: *IEEE intelligence vehicles symposium*, 106–113
22. Bhaskar H, Werghi N (2011) Comparing Hough and Radon transform based parallelogram detection. *GCC conference and exhibition (GCC)*, 641–644

Dynamic Surface Control of Hypersonic Aircraft with Parameter Estimation

Bin Xu, Fuchun Sun, Shixing Wang and Hao Wu

Abstract This paper investigates the adaptive controller for the longitudinal dynamics of a generic hypersonic aircraft. The control-oriented model is adopted for design. The subsystem is transformed into the linearly parameterized form. Based on the parameter projection estimation, the dynamic inverse control is proposed via back-stepping. The dynamic surface method is employed to provide the derivative information of the virtual control. The proposed methodology addresses the issue of controller design with respect to parametric model uncertainty. Simulation results show that the proposed approach achieves good tracking performance in the presence of uncertain parameters.

Keywords Hypersonic flight control · Dynamic surface control · Linearly parameterized form

1 Introduction

Hypersonic flight vehicles (HFVs) are intended to present a cost-efficient way to access space by reducing the flight time. The success of NASA's X-43A experimental airplane in flight testing has affirmed the feasibility of this technology. The U.S. military launched an experimental hypersonic aircraft on its swan song test

B. Xu (✉)

School of Automation, Northwestern Polytechnical University, Xi'an, China
e-mail: smileface.binxu@gmail.com

F. Sun · S. Wang

Department of Computer Science and Technology, Tsinghua University, Beijing, China

H. Wu

Department of Mathematics and Computer Science, Free University of Berlin, Berlin, Germany

flight on May 1, 2013, accelerating the craft to more than five times the speed of sound in the longest-ever mission for a vehicle of its kind.

The related hypersonic flight control has gained more and more attention. Based on linearizing the model at the trim state of the dynamics, the pivotal early works [1, 2] employed classic and multivariable linear control. The adaptive control [3] is investigated by linearizing the model at the trim state. Based on the input–output linearization using Lie derivative notation, the sliding mode control [4] is applied on Winged-Cone configuration [5]. The genetic algorithm [6] is employed for robust adaptive controller design.

In [7], the altitude subsystem is transformed into the strict-feedback form using the back-stepping scheme [8], the neural networks and Kriging system-based methods are investigated on discrete hypersonic flight control with nominal feedback [9–11]. The sequential loop closure controller design [12] is based on the equations decomposition into functional subsystems with the model from the assumed-modes version [13]. Based on locally valid linear-in-the-parameters non-linear model the unknown parameters are adapted by Lyapunov-based updating law. However, during the controller design, the back-stepping design needs repeated differentiations of the virtual control and it introduces more unknown items [14].

In this paper, the control-oriented model (COM) recently developed in [15] including the coupling effect of the engine to the airframe dynamics is studied. The subsystem is written into the linearly parameterized form. Instead of nominal feedback or fuzzy/neural approximation [16], the dynamic inverse control is proposed via back-stepping based on the parameter projection estimation. To avoid the “explosion of complexity” during the back-stepping design [12], the dynamic surface method is employed.

This paper is organized as follows. Section 2 briefly presents the COM of the generic HFV longitudinal dynamics. In Sect. 3, the dynamic inverse control is designed for the subsystems. The simulation is included in Sect. 4. Section 5 presents several comments and final remarks.

2 Hypersonic Vehicle Modeling

The control-oriented model of the longitudinal dynamics of a generic hypersonic aircraft from [15] is considered in this study. This model is comprised of five state variables $X_h = [V, h, \alpha, \gamma, q]^T$ and two control inputs $U_h = [\delta_e, \Phi]^T$.

$$\dot{V} = \frac{T \cos \alpha - D}{m} - g \sin \gamma \quad (1)$$

$$\dot{h} = V \sin \gamma \quad (2)$$

$$\dot{\gamma} = \frac{L + T \sin \alpha}{mV} - \frac{g \cos \gamma}{V} \quad (3)$$

$$\dot{\alpha} = q - \dot{\gamma} \quad (4)$$

$$\dot{q} = \frac{M_{yy}}{I_{yy}} \quad (5)$$

where

$$T \approx \bar{q}S \left(C_{T\Phi}^{\alpha^3} \alpha^3 + C_{T\Phi}^{\alpha^2} \alpha^2 + C_{T\Phi}^{\alpha} \alpha + C_{T\Phi}^0 \right) \Phi + \bar{q}S \left(C_T^{\alpha^3} \alpha^3 + C_T^{\alpha^2} \alpha^2 + C_T^{\alpha} \alpha + C_T^0 \right)$$

$$D \approx \bar{q}S \left(C_D^{\alpha^2} \alpha^2 + C_D^{\alpha} \alpha + C_D^0 \right)$$

$$L = L_0 + L_{\alpha} \alpha \approx \bar{q}S C_L^0 + \bar{q}S C_L^{\alpha} \alpha$$

$$M_{yy} = M_T + M_0(\alpha) + M_{\delta_e} \delta_e \approx z_T T + \bar{q}S \bar{c} \left(C_M^{\alpha^2} \alpha^2 + C_M^{\alpha} \alpha + C_M^0 \right) + \bar{q}S \bar{c} C_M^{\delta_e} \delta_e$$

$$\bar{q} = \frac{1}{2} \rho V^2, \rho = \rho_0 \exp \left[-\frac{h - h_0}{h_s} \right]$$

It is assumed that all of the coefficients of the model are subjected to uncertainty. The vector of all uncertain parameters, denoted by $p \in R^{L_p}$, includes the vehicle inertial parameters and the coefficients that appear in the force and moment approximations. The nominal value of p is denoted by p_0 . For simplicity, the maximum uniform variation within 30% of the nominal value has been considered, yielding the parameter set $\Omega_p = \{p \in R^{L_p} \mid p_i^L \leq p_i \leq p_i^U, i = 1, \dots, L_p\}$ and $p_i^L = \min \{0.7p_i^0, 1.3p_i^0\}$, $p_i^U = \max \{0.7p_i^0, 1.3p_i^0\}$.

3 Control Design

The control problem considered in this work takes into account only cruise trajectories and does not consider the ascent or the reentry of the vehicle. In the study [7, 9, 11], by functional decomposition, the velocity is independent with other subsystems. The goal pursued in this study is to design a dynamic controller Φ and δ_e to steer system velocity and altitude from a given set of initial values to desired trim conditions with the tracking reference V_r and h_r . Furthermore, the altitude command is transformed into the flight path angle (FPA) tracking. Define the altitude tracking error $\tilde{h} = h - h_r$. The demand of FPA is generated as

$$\gamma_d = \arcsin \left(\frac{-k_h \tilde{h} + \dot{h}_r}{V} \right) \quad (6)$$

where $k_h > 0$ is the design parameter.

3.1 Dynamic Inversion Control of Velocity Subsystem

Define the velocity error

$$\tilde{V} = V - V_r \quad (7)$$

From (7), the velocity dynamics are derived as

$$\dot{\tilde{V}} = \frac{T_\Phi \cos \alpha}{m} \Phi + \frac{T_0 \cos \alpha - D}{m} - g \sin \gamma - \dot{V}_r \quad (8)$$

Define $g_v = \frac{T_\Phi \cos \alpha}{m}$, $f_v = \frac{T_0 \cos \alpha - D}{m}$. Then Eq. (8) becomes

$$\dot{\tilde{V}} = g_v \Phi + f_v - g \sin \gamma - \dot{V}_r \quad (9)$$

where $f_v = \omega_{f_v}^T \theta_{f_v}$, $g_v = \omega_{g_v}^T \theta_{g_v}$ with

$$\begin{aligned} \omega_{f_v} &= \bar{q} S \begin{bmatrix} \alpha^3 \cos \alpha, \alpha^2 \cos \alpha, \alpha \cos \alpha, \cos \alpha, \\ -\alpha^2, -\alpha, -1 \end{bmatrix}^T \\ \theta_{f_v} &= \frac{1}{m} \begin{bmatrix} C_T^{\alpha^3}, C_T^{\alpha^2}, C_T^\alpha, C_T^0, C_D^{\alpha^2}, C_D^\alpha, \\ C_D^0 \end{bmatrix}^T \\ \omega_{g_v} &= \bar{q} S \begin{bmatrix} \alpha^3 \cos \alpha, \alpha^2 \cos \alpha, \alpha \cos \alpha, \cos \alpha \end{bmatrix}^T \\ \theta_{g_v} &= \frac{1}{m} \begin{bmatrix} C_{T\Phi}^{\alpha^3}, C_{T\Phi}^{\alpha^2}, C_{T\Phi}^\alpha, C_{T\Phi}^0 \end{bmatrix}^T \end{aligned}$$

The throttle setting is designed as

$$\hat{g}_v \Phi = -k_v \tilde{V} - \hat{f}_v + g \sin \gamma + \dot{V}_r \quad (10)$$

where $k_v > 0$ is a design parameter, $\hat{f}_v = \omega_{f_v}^T \hat{\theta}_{f_v}$ and $\hat{g}_v = \omega_{g_v}^T \hat{\theta}_{g_v}$.

Then Eq. (8) can be expressed as

$$\dot{\tilde{V}} = \tilde{g}_v \Phi + \tilde{f}_v - k_v \tilde{V} \quad (11)$$

where $\tilde{f}_v = \omega_{f_v}^T (\theta_{f_v} - \hat{\theta}_{f_v}) = \omega_{f_v}^T \tilde{\theta}_{f_v}$, $\tilde{g}_v = \omega_{g_v}^T (\theta_{g_v} - \hat{\theta}_{g_v}) = \omega_{g_v}^T \tilde{\theta}_{g_v}$.

The control Lyapunov function candidate for the velocity error dynamics is selected as

$$W_V = \frac{1}{2} \left(\tilde{V}^2 + \tilde{\theta}_{f_v}^T \Gamma_{f_v}^{-1} \tilde{\theta}_{f_v} + \tilde{\theta}_{g_v}^T \Gamma_{g_v}^{-1} \tilde{\theta}_{g_v} \right) \quad (12)$$

The derivative of W_V is

$$\begin{aligned} \dot{W}_V &= \tilde{V} \dot{\tilde{V}} - \tilde{\theta}_{f_v}^T \Gamma_{f_v}^{-1} \dot{\tilde{\theta}}_{f_v} - \tilde{\theta}_{g_v}^T \Gamma_{g_v}^{-1} \dot{\tilde{\theta}}_{g_v} \\ &= -k_v \tilde{V}^2 + \tilde{V} \omega_{f_v}^T \tilde{\theta}_{f_v} + \tilde{V} \omega_{g_v}^T \tilde{\theta}_{g_v} \Phi - \tilde{\theta}_{f_v}^T \Gamma_{f_v}^{-1} \dot{\tilde{\theta}}_{f_v} - \tilde{\theta}_{g_v}^T \Gamma_{g_v}^{-1} \dot{\tilde{\theta}}_{g_v} \\ &= -k_v \tilde{V}^2 - \tilde{\theta}_{f_v}^T \left(\Gamma_{f_v}^{-1} \dot{\tilde{\theta}}_{f_v} - \tilde{V} \omega_{f_v} \right) - \tilde{\theta}_{g_v}^T \left(\Gamma_{g_v}^{-1} \dot{\tilde{\theta}}_{g_v} - \tilde{V} \omega_{g_v} \Phi \right) \end{aligned} \quad (13)$$

The adaptive law is designed as

$$\dot{\hat{\theta}}_{fv} = \text{Proj} (\Gamma_{fv} \tilde{V} \omega_{fv}) \quad (14)$$

$$\dot{\hat{\theta}}_{gv} = \text{Proj} (\Gamma_{gv} \tilde{V} \omega_{gv} \Phi) \quad (15)$$

Then

$$\dot{W}_V = -k_v \tilde{V}^2 \quad (16)$$

It is easy to know that the velocity is asymptotically stable.

3.2 Dynamic Surface Control of Attitude Subsystem

Define $x_1 = \gamma, x_2 = \theta_p, x_3 = q, \theta_p = \alpha + \gamma, u = \delta_e$. The following subsystem can be obtained

$$\begin{aligned} \dot{x}_1 &= g_1 x_2 + f_1 - \frac{g}{V} \cos x_1 \\ \dot{x}_2 &= x_3 \\ \dot{x}_3 &= g_3 u + f_3 \end{aligned} \quad (17)$$

where

$$\begin{aligned} f_1 &= \frac{L_0 - L_\alpha \gamma + T \sin \alpha}{mV} = \omega_{f1}^T \theta_{f1} \\ g_1 &= \frac{L_\alpha}{mV} = \omega_{g1}^T \theta_{g1} \\ f_3 &= \frac{M_T + M_0(\alpha)}{I_{yy}} = \omega_{f3}^T \theta_{f3} \\ g_3 &= \frac{M_{\delta_e}}{I_{yy}} = \omega_{g3}^T \theta_{g3} \end{aligned}$$

with

$$\begin{aligned} \omega_{f1} &= \frac{\bar{q}S}{V} \begin{bmatrix} 1, -\gamma, \alpha^3 \Phi \sin \alpha, \alpha^2 \Phi \sin \alpha, \alpha \Phi \sin \alpha, \Phi \sin \alpha, \alpha^3 \sin \alpha, \alpha^2 \sin \alpha, \alpha \sin \alpha, \sin \alpha \end{bmatrix}^T \\ \theta_{f1} &= \frac{1}{m} \begin{bmatrix} C_L^0, C_L^\alpha, C_{T\Phi}^{\alpha^3}, C_{T\Phi}^{\alpha^2}, C_{T\Phi}^\alpha, C_{T\Phi}^0, \\ C_T^{\alpha^3}, C_T^{\alpha^2}, C_T^\alpha, C_T^0 \end{bmatrix}^T \end{aligned}$$

$$\begin{aligned}\omega_{g1} &= \frac{\bar{q}S}{V}, \theta_{g1} = \frac{1}{m} C_L^\alpha \\ \omega_{f3} &= \bar{q}S [\alpha^3 \Phi, \alpha^2 \Phi, \alpha \Phi, \Phi, \alpha^3, \alpha^2, \alpha, 1, \alpha^2, \alpha, 1]^T \\ \theta_{f3} &= \frac{1}{I_{yy}} \begin{bmatrix} z_T \left(C_{T\Phi}^{\alpha^3}, C_{T\Phi}^{\alpha^2}, C_{T\Phi}^\alpha, C_{T\Phi}^0 \right), \\ z_T \left(C_T^{\alpha^3}, C_T^{\alpha^2}, C_T^\alpha, C_T^0 \right), \\ \bar{c} \left(C_M^{\alpha^2}, C_M^\alpha, C_M^0 \right) \end{bmatrix}^T \\ \omega_{g3} &= \bar{q}S, \theta_{g3} = \frac{1}{I_{yy}} \bar{c} C_M^{\delta_e}\end{aligned}$$

Step 1. Define $\tilde{x}_1 = x_1 - x_{1d}$. The dynamics of the flight path angle tracking error \tilde{x}_1 are written as

$$\dot{\tilde{x}}_1 = \dot{x}_1 - \dot{x}_{1d} = g_1 x_2 + f_1 - \frac{g}{V} \cos \gamma - \dot{x}_{1d} \quad (18)$$

Take θ_p as virtual control and design x_{2c} as

$$\hat{g}_1 x_{2c} = -k_1 \tilde{x}_1 - \hat{f}_1 + \frac{g}{V} \cos x_1 + \dot{x}_{1d} \quad (19)$$

where $k_1 > 0$ is the design parameter, $\hat{f}_1 = \omega_{f1}^T \hat{\theta}_{f1}$, $\hat{g}_1 = \omega_{g1}^T \hat{\theta}_{g1}$. Introduce a new state variable x_{2d} , which can be obtained by the following first-order filter

$$\varepsilon_2 \dot{x}_{2d} + x_{2d} = x_{2c}, x_{2d}(0) = x_{2c}(0) \quad (20)$$

Define $y_2 = x_{2d} - x_{2c}$, $\tilde{x}_2 = x_2 - x_{2d}$.

$$\begin{aligned}\dot{\tilde{x}}_1 &= g_1 x_2 + f_1 - \frac{g}{V} \cos \gamma - \dot{x}_{1d} \\ &= g_1 (x_2 - x_{2c}) + g_1 x_{2c} - \hat{g}_1 x_{2c} + \hat{g}_1 x_{2c} + f_1 - \frac{g}{V} \cos \gamma - \dot{x}_{1d} \\ &= g_1 (x_2 - x_{2c}) + \tilde{g}_1 x_{2c} + \tilde{f}_1 - k_1 \tilde{x}_1 \\ &= g_1 \tilde{x}_2 + g_1 y_2 + \tilde{g}_1 x_{2c} + \tilde{f}_1 - k_1 \tilde{x}_1\end{aligned} \quad (21)$$

The adaption laws of the estimated parameters are

$$\dot{\hat{\theta}}_{f1} = \text{Proj} (\Gamma_{f1} \omega_{f1} \tilde{x}_1) \quad (22)$$

$$\dot{\hat{\theta}}_{g1} = \text{Proj} (\Gamma_{g1} \omega_{g1} \tilde{x}_1 x_{2c}) \quad (23)$$

Step 2. The dynamics of the pitch angle tracking error \tilde{x}_2 are written as

$$\dot{\tilde{x}}_2 = \dot{x}_2 - \dot{x}_{2d} = x_3 - \dot{x}_{2d} \quad (24)$$

Take q as virtual control and design x_{3c} as

$$x_{3c} = -k_2\tilde{x}_2 + \dot{x}_{2d} \quad (25)$$

where $k_2 > 0$ is the design parameter.

Introduce a new state variable x_{3d} , which can be obtained by the following first-order filter

$$\varepsilon_3\dot{x}_{3d} + x_{3d} = x_{3c}, x_{3d}(0) = x_{3c}(0) \quad (26)$$

Define $y_3 = x_{3d} - x_{3c}$, $\tilde{x}_3 = x_3 - x_{3d}$.

$$\begin{aligned} \dot{\tilde{x}}_2 &= x_3 - \dot{x}_{2d} \\ &= x_3 - x_{3d} + x_{3d} - x_{3c} + x_{3c} - \dot{x}_{2d} \\ &= \tilde{x}_3 + y_3 - k_2\tilde{x}_2 \end{aligned} \quad (27)$$

Step 3. The dynamics of the pitch rate tracking error \tilde{x}_3 are written as

$$\dot{\tilde{x}}_3 = \dot{x}_3 - \dot{x}_{3d} = g_3u + f_3 - \dot{x}_{3d} \quad (28)$$

Design the elevator deflection δ_e as

$$\hat{g}_3u = -k_3\tilde{x}_3 - \hat{f}_3 + \dot{x}_{3d} \quad (29)$$

where $k_3 > 0$ is the design parameter, $\hat{f}_3 = \omega_{f_3}^T \hat{\theta}_{f_3}$, $\hat{g}_3 = \omega_{g_3}^T \hat{\theta}_{g_3}$.

The error dynamics are derived as

$$\begin{aligned} \dot{\tilde{x}}_3 &= g_3u + f_3 - \dot{x}_{3d} \\ &= (\tilde{g}_3 + \hat{g}_3)u + f_3 - \dot{x}_{3d} \\ &= \tilde{g}_3u - k_3\tilde{x}_3 + \tilde{f}_3 \end{aligned} \quad (30)$$

The adaption laws of the estimated parameters are

$$\dot{\hat{\theta}}_{f_3} = \text{Proj} (\Gamma_{f_3} \omega_{f_3} \tilde{x}_3) \quad (31)$$

$$\dot{\hat{\theta}}_{g_3} = \text{Proj} (\Gamma_{g_3} \omega_{g_3} \tilde{x}_3 u) \quad (32)$$

Assumption 1 The FPA reference signal and its derivatives are smooth bounded functions.

Assumption 2 There exists constant $\bar{g}_1 > |g_1| > 0$.

Select Lypunov function

$$W = \sum_{i=1}^3 W_i \quad (33)$$

with

$$\begin{aligned}
 W_1 &= \frac{1}{2} \left(\tilde{x}_1^2 + \tilde{\theta}_{f1}^T \Gamma_{f1}^{-1} \tilde{\theta}_{f1} + \tilde{\theta}_{g1}^T \Gamma_{g1}^{-1} \tilde{\theta}_{g1} + y_2^2 \right) \\
 W_2 &= \frac{1}{2} (\tilde{x}_2^2 + y_3^2) \\
 W_3 &= \frac{1}{2} \left(\tilde{x}_3^2 + \tilde{\theta}_{f3}^T \Gamma_{f3}^{-1} \tilde{\theta}_{f3} + \tilde{\theta}_{g3}^T \Gamma_{g3}^{-1} \tilde{\theta}_{g3} \right)
 \end{aligned}$$

Theorem 1. Consider system (17) with virtual control (19), (25), actual control (29) with adaption laws (22), (23), (31) and (32) under Assumptions 1–2. Then all the signals of (33) are uniformly ultimately bounded.

Remark for each W_i , one can follow the analysis procedure in velocity subsystem. The proof could be done by following the procedure in [14] and thus it is omitted here. The work was part of the design and analysis of the DSC based actuator saturation control [17].

4 Simulations

The rigid body of the hypersonic flight vehicle is considered in the simulation study. The parameters for COM can be found in [15]. The reference commands are generated by the filter

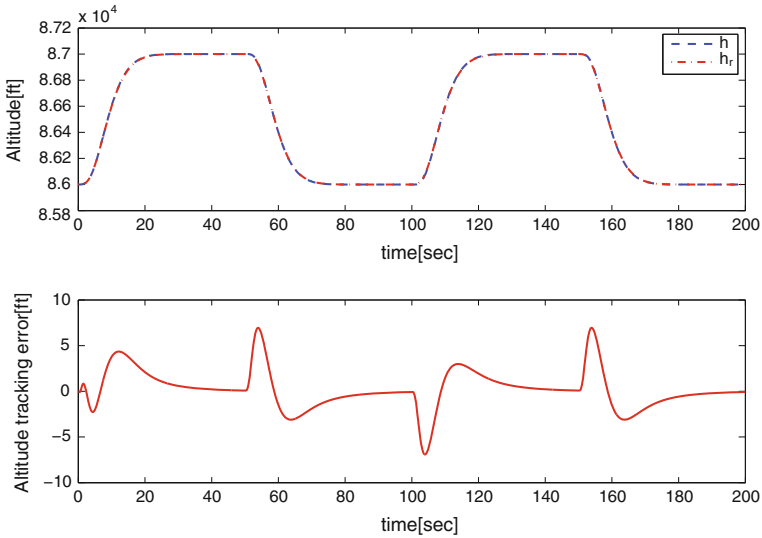


Fig. 1 Altitude tracking

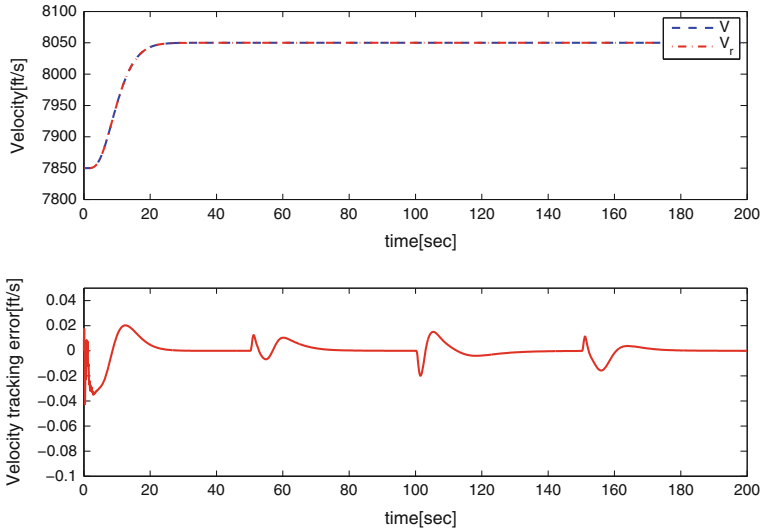
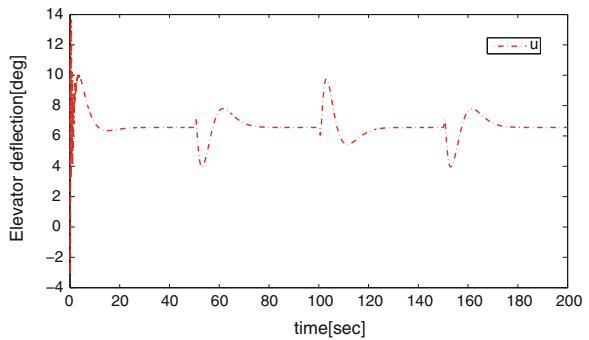


Fig. 2 Velocity tracking

Fig. 3 Elevator deflection



$$\frac{h_r}{h_c} = \frac{0.16}{(s^2 + 0.76s + 0.16)} \tag{34}$$

$$\frac{V_r}{V_c} = \frac{0.16}{(s^2 + 0.76s + 0.16)} \tag{35}$$

The control gains for the dynamic surface controller are selected as [6, 0.3, 2, 5, 8] separately for $[k_v, k_h, k_1, k_2, k_3]$, and the first-order filter parameter for dynamic surface design is $\varepsilon_i = 0.02, i = 2, 3$. Parameters for projection algorithm are selected as $\Gamma_{fi} = 0.1I, \Gamma_{gi} = 0.1I, i = 1, 3, v$.

The initial values of the states are set as $v_0 = 7,850$ ft/s, $h_0 = 86,000$ ft, $\alpha_0 = 3.5^\circ, \gamma_0 = 0, q_0 = 0$. The velocity tracks the step command with 200ft/s while the altitude follows the square command with period 100 s and magnitude 1,000 ft.

Fig. 4 Throttle setting

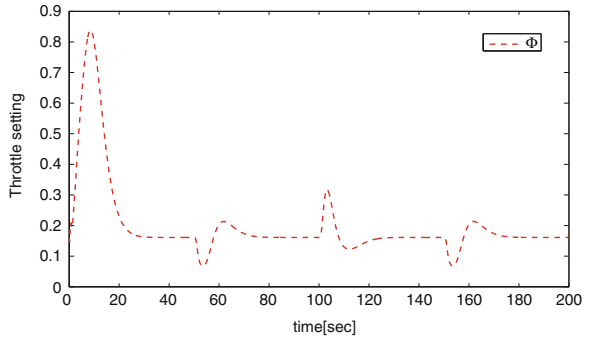
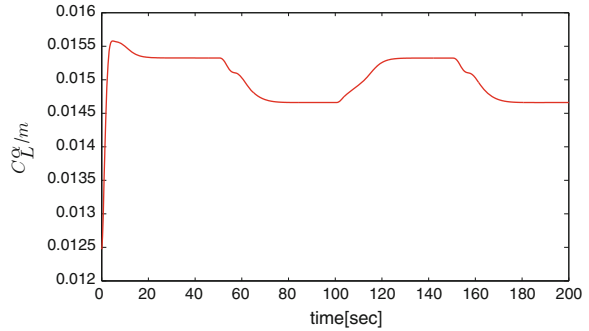


Fig. 5 Estimation of C_L^z/m



The satisfied tracking performance is depicted in Figs. 1 and 2.

The altitude follows the square signal while the velocity is responding to the step command. From the control input referred to

first period is larger than others. This is due to the fact that velocity is stepped from 7,850 to 8,050 ft/s in about 20 s and it is kept stable in the next periods with small variation which is caused from the square tracking of the altitude. The elevator deflection is changing fast at the beginning. The reason could be found in the parameter estimation in Fig. 5 where the estimation responds in a large domain and then later it is stable. The simulation shows the robustness of the algorithm regarding to the parameter uncertainty (Fig. 3).

5 Conclusions and Future Work

The dynamics of HFV are transformed into the linearly parameterized form. To avoid the “explosion of complexity,” the dynamic surface control is investigated on HFV. The closed-loop system achieves uniformly ultimately bounded stability. The effectiveness is verified by simulation study with parametric model

uncertainty. For future work, we will focus on the design in the presence of flexible states.

Acknowledgments This work was supported by the DSO National Laboratories of Singapore through a Strategic Project Grant (Project No: DSOCL10004), National Science Foundation of China (Grant No:61134004), NWPU Basic Research Funding (Grant No: JC20120236), and Deutsche Forschungsgemeinschaft (DFG) Grant No. WU 744/1-1.

References

1. Schmidt D (1992) Dynamics and control of hypersonic aeropropulsive/aeroelastic vehicles. AIAA Paper, pp 1992-4326
2. Schmidt D (1997) Optimum mission performance and multivariable flight guidance for airbreathing launch vehicles. *J Guidance Control Dyn* 20(6):1157-1164.
3. Gibson T, Crespo L, Annaswamy A (2009) Adaptive control of hypersonic vehicles in the presence of modeling uncertainties. American Control Conference, Missouri, USA, June, pp 3178-3183
4. Xu H, Mirmirani M, Ioannou P (2004) Adaptive sliding mode control design for a hypersonic flight vehicle. *J Guidance Control Dyn* 27(5):829-838
5. Shaughnessy J, Pinckney S, McMinn J, Cruz C, Kelley M (1990) Hypersonic vehicle simulation model: Winged-Cone configuration. NASA TM 102610, Nov 1990
6. Wang Q, Stengel R (2000) Robust nonlinear control of a hypersonic aircraft. *J Guidance Control Dyn* 23(4):577-585
7. Gao DX, Sun ZQ (2011) Fuzzy tracking control design for hypersonic vehicles via TS model. *Sci China Inf Sci* 54(3):521-528
8. Kokotovic P (1991) The joy of feedback: nonlinear and adaptive: 1991 bode prize lecture. *IEEE Control Syst Mag* 12:7-17
9. Xu B, Sun F, Yang C, Gao D, Ren J (2011) Adaptive discrete-time controller design with neural network for hypersonic flight vehicle via back-stepping. *Int J Control* 84(9): 1543-1552
10. Xu B, Sun F, Liu H, Ren J (2012) Adaptive Kriging controller design for hypersonic flight vehicle via back-stepping. *IET Control Theory Appl* 6(4):487-497
11. Xu B, Wang D, Sun F, Shi Z (2012) Direct neural discrete control of hypersonic flight vehicle. *Nonlinear Dyn* 70(1):269-278
12. Fiorentini L, Serrani A, Bolender M, Doman D (2008) Robust nonlinear sequential loop closure control design for an air-breathing hypersonic vehicle model. American Control Conference, Seattle, USA, pp 3458-3463
13. Williams T, Bolender M, Doman D, Morataya O (2006) An aerothermal flexible mode analysis of a hypersonic vehicle. In: AIAA Atmospheric Flight Mechanics Conference and Exhibit, Keystone, AIAA Paper, pp 2006-6647.
14. Wang D, Huang J (2005) Neural network-based adaptive dynamic surface control for a class of uncertain nonlinear systems in strict-feedback form. *IEEE Trans Neural Networks* 16(1):195-202
15. Parker J, Serrani A, Yurkovich S, Bolender M, Doman D (2007) Control-oriented modeling of an air-breathing hypersonic vehicle. *J Guidance Control Dyn* 30(3):856-869
16. Xu B, Gao D, Wang S (2011) Adaptive neural control based on HGO for hypersonic flight vehicles. *Sci China Inf Sci* 54(3):511-520
17. Xu B, Huang X, Wang D, Sun F. Dynamic surface control of constrained hypersonic flight models with parameter estimation and actuator compensation. *Asian J Control*. doi:10.1002/asjc.679

Nonlinear Optimal Control for Robot Manipulator Trajectory Tracking

Shijie Zhang, Ning Yi and Fengzhi Huang

Abstract This paper presents a nonlinear optimal feedback control approach for robot manipulators with dynamics nonlinearities. The task of tracking a preplanned trajectory of robot manipulator is formulated as an optimal control problem, in which the energy consumption and motion time are minimized. The optimal control problem is first solved as an open-loop optimal control problem by using a time-scaling transform and the control parameterization method. Then, by virtue of the relationship between the optimal open-loop control and the optimal closed-loop control along the optimal trajectory, a practical method is presented to calculate an approximate optimal feedback gain matrix, without having to solve an optimal control problem involving the complex Riccati-like matrix differential equation coupled with the original system dynamics. Simulation results of two-link robot manipulator are presented to show that the proposed approach is highly effective.

Keywords Nonlinear control · Optimal · Robot manipulator · Trajectory tracking

1 Introduction

A basic problem in controlling robots is to make the manipulator to follow a desired trajectory. High-speed and high-precision trajectory trackings are frequently requirements for applications of robot manipulators.

S. Zhang (✉) · N. Yi · F. Huang
Research Institute of Robotics, Henan University of Technology, Zhengzhou, China
e-mail: zhangshijie@haut.edn.cn

N. Yi
e-mail: robot@haut.edn.cn

F. Huang
e-mail: huangfengzhi@haut.edn.cn

The optimal control schemes for manipulator arms have been actively researched in robotics in the past two decades because the optimal motions that minimize energy consumption, error trajectories, or motion time yield high productivity, efficiency, smooth motion, durability of machine parts, etc. [1–6]. Various types of methods have been developed to solve the robotic manipulator optimal control schemes. By the application of the optimal control theory, Pontryagin's maximum principle leads to a two-point boundary value problem. Although this theory and its solutions are rigorous, it has been used to solve equations for the motions of two-link- or at most three-link-planar manipulators due to the complexity and the nonlinearity of the manipulator dynamics [1]. Approximation methods have been studied to obtain the solutions for three or more DOF spatial manipulators. However, the solutions obtained have not been proved to be optimal. These approximation methods are roughly divided into two groups depending on whether or not they utilize gradients [5]. The method in paper [7] is to represent the nonlinear system into a sequence of linear time-varying system through the recursive theory of approximation. However, the linearized model does not compensate the nonlinearities in the system due to the wide range of operating conditions. Recently, the applications of intelligent control techniques (such as fuzzy control or neural network control) with optimal algorithm to the motion control for robot manipulators have received considerable attention [8–13]. But sometimes, these methods take quite a long time to find a coefficient that satisfies the requirement of the controlling task. In addition, lack of theoretical analysis and stability security makes industrialists wary of using the results in real industrial environments. This paper is concerned with the nonlinear optimal feedback control for robot manipulator trajectory tracking. The energy consumption and error trajectories are minimized as performance index in the optimal control problem. An optimal open-loop control is first obtained by using a time-scaling transform [14] and the control parameterization technique [15]. Then, we derive the form of the optimal closed-loop control law, which involves a feedback gain matrix, for the optimal control problem. The optimal feedback gain matrix is required to satisfy a Riccati-like matrix differential equation. Then, the third-order B-spline function, which has been proved to be very efficient for solving optimal approximation and optimal control problems, is employed to construct the components of the feedback gain matrix. By virtue of the relationship between the optimal open-loop control and the optimal closed-loop control along the optimal trajectory, a practical computational method is presented for finding an approximate optimal feedback gain matrix, without having to solve an optimal control problem involving the complex Riccati-like matrix differential equation coupled with the original system dynamics [16].

2 Nonlinear Dynamic of Robot Manipulator

2.1 Models of Robot Dynamics

Consider the dynamic equation of a robot manipulator

$$M(q)\ddot{q} + C(q, \dot{q})\dot{q} + g(q) = u(t) \tag{1}$$

where $q, \dot{q}, \ddot{q} \in \mathbb{R}^n$ are the vectors of the generalized joint coordinates, velocity, and acceleration, $M(q) \in \mathbb{R}^{n \times n}$ denotes a symmetric positive definite inertia matrix, $C(q, \dot{q}) \in \mathbb{R}^{n \times n}$ stands for the Coriolis and centrifugal torques, $g(q) \in \mathbb{R}^n$ models the gravity forces, and $u(t) \in \mathbb{R}^n$ is the torque input. Some useful properties of robot dynamic are as follows:

Property 1: Matrix $M(q)$ is symmetric and positive definite.

Property 2: Matrix $\dot{M}(q) - 2C(q, \dot{q})$ is skew symmetric and satisfies that

$$\dot{q}^T [\dot{M}(q) - 2C(q, \dot{q})] \dot{q} = 0$$

Property 3: The robot dynamics is passive in open-loop, from torque input to velocity output, with the Hamiltonian as its storage function. If viscous friction was considered, the energy dissipates and the system is strictly passive.

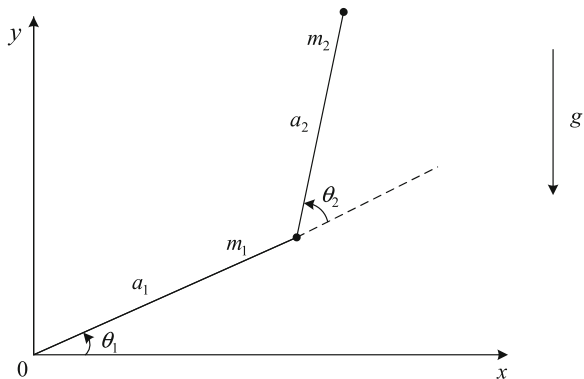
The two-link-revolute robot (RR) manipulator is shown in Fig. 1. The masses of both links and actuators are denoted by m_1, m_2 with I_1, I_2 as mass moment of inertia. a_1, a_2 denotes the length, u_1, u_2 are joints torques. The joints positions of the two links are defined by θ_1, θ_2 .

The dynamic equations of two-link RR manipulator are written in state space form as

$$\dot{x} = f(x) + B(x)u(t) \tag{2}$$

where $x = [q^T, \dot{q}^T]^T$ is the system state, $q = [\theta_1, \theta_2]^T$, and

Fig. 1 Two-link RR manipulator



$$f(x) = \begin{bmatrix} \dot{q} \\ -M^{-1}(q)(C(q, \dot{q})\dot{q} + g(q)) \end{bmatrix}$$

$$B(x) = \begin{bmatrix} 0 \\ M^{-1}(q) \end{bmatrix}$$

where

$$M(q) = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

$$c_{11} = (m_1 + m_2)a_1^2 + m_2a_2^2 + 2m_2a_1a_2 \cos \theta_2$$

$$c_{12} = c_{21} = m_2a_2^2 + m_2a_1a_2 \cos \theta_2$$

$$c_{22} = m_2a_2^2$$

$$C(q, \dot{q}) = \begin{bmatrix} -m_2a_1a_2(2\dot{\theta}_1\dot{\theta}_2 + \dot{\theta}_2^2) \sin \theta_2 \\ m_2a_1a_2\dot{\theta}_1^2 \sin \theta_2 \end{bmatrix}$$

$$g(q) = \begin{bmatrix} (m_1 + m_2)ga_1 \cos \theta_1 + m_2ga_2 \cos(\theta_1 + \theta_2) \\ m_2ga_2 \cos(\theta_1 + \theta_2) \end{bmatrix}$$

Define

$$N(q, \dot{q}) = C(q, \dot{q}) + g(q) = \begin{bmatrix} N_1(q, \dot{q}) \\ N_2(q, \dot{q}) \end{bmatrix}$$

and

$$M^{-1}(q) = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$

then,

$$f(x) = \begin{bmatrix} \theta_2 \\ \dot{\theta}_2 \\ \Theta\dot{\theta}_1 \\ \Xi\theta_1 \end{bmatrix}, B(x) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$

here

$$\Theta = \frac{M_{22}(-N_2(q, \dot{q})) + M_{12}(-N_1(q, \dot{q}))}{x_1(t)}$$

$$\Xi = \frac{M_{22}(-N_2(q, \dot{q})) + M_{11}(-N_1(q, \dot{q}))}{x_2(t)}.$$

2.2 Problem Statement

The purpose of control is to determine an optimal closed-loop control signal so that the robot manipulator tracks the desired trajectory with minimal energy consumption, The optimal control problem can be formulated as follows.

Given system (2), find a closed-loop control $u(t) \in \mathbb{R}^n$ such that the cost function

$$J = \alpha_1 \Phi_0(x(T)) + \alpha_2 \int_0^T u^T R u dt \tag{3}$$

is minimized, where $\Phi_0(x(T)) = (x(T) - x_d)^T Q (x(T) - x_d)$, T is the free terminal time, x_d is the desired trajectory, α_1 and α_2 are the weighting parameters, $Q \in \mathbb{R}^{2n \times 2n}$ and $R \in \mathbb{R}^{n \times n}$ are symmetric positive semi-definite and symmetric positive definite weighting matrices, respectively.

We refer to the above problem as Problem (P). This optimal close-loop control problem is very difficult to solve directly. In this paper, we derive the form of the optimal closed-loop control law after obtaining an optimal open-loop control by using a time-scaling transform and the control parameterization technique. Then the difficult of the problem is transformed to find a feedback gain matrix which is involved in the optimal closed-loop control law. A practical computational method is presented in [16] for finding an approximate optimal feedback gain matrix, without having to solve an optimal control problem involving the complex Riccati-like matrix differential equation coupled with the original system dynamics.

3 Nonlinear Optimal Control Design

By using a time-scaling transform and the control parameterization technique, the above problem is solved as an optimal open-loop control problem firstly. An optimal open-loop control and the corresponding optimal trajectory will be provided.

Let the time horizon $[0, T]$ be partitioned into p subintervals as follows:

$$0 = t_0 \leq t_1 \leq \dots \leq t_p = T. \tag{4}$$

The switching times $t_i, 1 \leq i \leq p$, are regarded as decision variables. Employing the time-scaling transform introduced in [14] to map these switching times into a set of fixed time points $\theta_i = \frac{i}{p}, i = 1, \dots, p$, on a new time horizon $[0, 1]$. Then the following differential equation is achieved

$$\frac{dt(s)}{ds} = v^p(s), s \in [0, 1] \tag{5}$$

where

$$v^p(s) = \sum_{i=1}^p \xi_i \chi_{[\theta_{i-1}, \theta_i]}(s) \tag{6}$$

where $\chi_I(s)$ denotes the indicator function of I defined by

$$\chi_I(s) = \begin{cases} 1, & s \in I \\ 0, & \text{elsewhere} \end{cases} \tag{7}$$

and $\xi_i \geq 0, \sum_{i=1}^p \xi_i = T$.

For $s \in [\theta_{l-1}, \theta_l]$, we have

$$t(s) = \sum_{i=1}^{l-1} \xi_i + \xi_l(s - \theta_{l-1})p, \tag{8}$$

where $l = 1, \dots, p$. Clearly,

$$t(1) = \sum_{i=1}^p \xi_i = T. \tag{9}$$

Then after the time-scaling transform, system (2) can be converted into the following form

$$\hat{x}(s) = v^p(s)[f(\hat{x}(s)) + B(\hat{x}(s), s)\tilde{u}(s)] \tag{10}$$

where $\hat{x}(s) = [\tilde{x}(s)^T, t(s)]^T, \tilde{x}(s) = x(t(s))$ and $\tilde{u}(s) = u(t(s))$.

Now we apply the control parameterization technique to approximate the control $\tilde{u}(s)$ as follows:

$$\tilde{u}_i^p(s) = \sum_{k=-1}^{p+1} \sigma_k^i \Omega\left(\left(\frac{1}{p}\right)s - k\right), i = 1, \dots, n \tag{11}$$

where

$$\Omega(\kappa) = \begin{cases} 0, & |\kappa| > 2 \\ -\frac{1}{6}|\kappa|^3 + \kappa^2 - 2|\kappa| + \frac{4}{3}, & 1 \leq |\kappa| \leq 2 \\ \frac{1}{2}|\kappa|^3 - \kappa^2 + \frac{2}{3}, & |\kappa| < 2 \end{cases} \tag{12}$$

is the cubic spline basis function.

Define $\sigma^i = [\sigma_{-1}^i, \dots, \sigma_{p+1}^i]^T, i = 1, \dots, n$, and $\sigma = [(\sigma^1)^T, \dots, (\sigma^n)^T]^T$, let Π denotes the set containing all σ . Then $\tilde{u}^p(s) = [\tilde{u}_1^p(s), \dots, \tilde{u}_n^p(s)]^T$ is determined uniquely by the switching vector $\sigma \in \Pi$. Thus, it can be written as $\tilde{u}^p(\cdot|\sigma)$. Now the optimal parameterization selection problem, which is an approximation of Problem (P), can be stated as follows:

Problem (Q). Given system (10), find a combined vector (σ, ξ) , such that the cost function

$$J(\sigma) = \alpha_1 \hat{\Phi}_0(\hat{x}(1|\sigma)) + \alpha_2 \int_0^1 v^p(s|\xi) \tilde{u}^p(s|\sigma)^T R \tilde{u}^p(s|\sigma) ds$$

is minimized, where $\hat{\Phi}_0(\hat{x}(1|\sigma)) = (\hat{x}(1|\sigma) - \hat{x}_d)^T \hat{S}(\hat{x}(1|\sigma) - \hat{x}_d)$, \hat{x}_d is the desired trajectory.

Now Problem (P) is approximated by a sequence of optimal parameter selection problems, each of which can be viewed as a mathematical programming problem and hence can be solved by existing gradient-based optimization methods. Here, our controls are approximated in terms of cubic spline basis functions, and thus they are smooth. Problem (Q) can be solved easily by use of the optimal control software package MISER3.3 [17].

Suppose that $(\tilde{u}^{p*}, \hat{x}^*)$ is the optimal solution of Problem (Q). Then it follows that the optimal solution to Problem (P) is (u^*, x^*, T^*) , where u^* is the optimal open-loop control, x^* is the corresponding optimal state vector, and T^* is the optimal terminal time. For the computation of the optimal closed-loop control problem, we have the following theorem.

Theorem 1. *The optimal closed-loop control \bar{u}^* for Problem (P) is given by*

$$\bar{u}^*(t) = \frac{1}{2\alpha_2} R^{-1} B^T K(t) f(x^*(t), t) \tag{13}$$

where x^* is the optimal state, and $K(t)$ is the solution of the following Riccati-like differential equation

$$\left(\dot{K} + KF + F^T K + \frac{1}{2} KFBR^{-1}B^TK \right) f + KD = 0 \tag{14}$$

where

$$F = \frac{\partial f}{\partial x}, D = \frac{\partial f}{\partial t}$$

and

$$K(T)f(x(T), T) = \alpha_1 \frac{\partial \Phi_0(x(T))}{\partial x(T)} = 2\alpha_1(x(T) - x_d)S.$$

The proof is similar to that given for Theorem 3.1 in [18]. Details can reference this literature.

By Theorem 1, Although the form of the optimal closed-loop control law is given, the matrix function $K(t)$ is still required to be obtained. The solving process involves solving a new optimal control problem denoted as follows. Using the method proposed in [16], Problem (R) could be solved well.

Problem (R). Subject to the dynamical system (1), with \bar{u} given by Theorem 1, find a $K(t)$ such that the cost function (11) also with \bar{u} is minimized.

In [16], an alternative approach was proposed to construct an approximate optimal matrix function $K^*(t)$ without having to solve this complicated optimal control problem (R). The basic idea is explained as follows. Suppose that u^* is an optimal open-loop control of Problem (P) and that x^* is the corresponding optimal state. We now consider Problem (P) with $x = x^*$, i.e. along the optimal open-loop path, and our task is to find a $K^*(t)$ such that $\check{u}^* = \frac{1}{2\alpha_2} R^{-1} B^T K^*(t) f(x^*(t), t)$ best approximates the control \bar{u}^* in the mean-square sense. Then \check{u}^* can be regarded as a good approximate optimal feedback control for Problem (P).

The calculation steps of solving $K^*(t)$ are as follows:

Step 1. The time horizon $[0, T^*]$ is partitioned into p equal subintervals,

$$0 = t_0 \leq t_1 \leq \dots \leq t_p \leq t_{p+1} = T^* \tag{15}$$

Step 2. Let

$$\left[K(t)_{i,j} \right] \approx \sum_{k=-1}^{p+1} (c_{i,j,k}) \Omega \left(\left(\frac{T^*}{p} \right) t - k \right) \tag{16}$$

where $c_{i,j,k}, i, j = 1, 2, \dots, n$ and $k = -1, 0, \dots, p + 1$, are real constant coefficients that are to be determined. p is the number of equality subintervals on $[0, T^*]$, $p + 3$ is the total number of cubic spline basis functions used in the approximation of each $\left[K(t)_{i,j} \right]$.

Step 3. Let

$$\mathcal{Y}(K) = \int_0^{T^*} \|u^*(t) - \check{u}(t)\|^2 dt \tag{17}$$

where

$$\check{u}(t) = \frac{1}{2\alpha_2} R^{-1} B^T K(t) f(x^*(t), t)$$

Step 4. Find coefficients $c_{i,j,k}$ such that the cost function (17) is minimized. These optimal coefficients can be obtained by solving the following optimality conditions

$$A = \frac{\partial \mathcal{Y}(K)}{\partial c_{i,j,k}} = 0 \tag{18}$$

we can see that these are linear equations and hence are easy to solve.

4 Simulation

In this section, the simulations of the nonlinear optimal control for the two-link RR manipulator are performed to show the efficiency of the proposed method.

Assuming that the friction is negligible. Two-link RR manipulators are simulated with following parameters.

$$m_1 = 1 \text{ Kg}, m_2 = 1 \text{ Kg}, a_1 = 1 \text{ m}, a_2 = 1 \text{ m}, g = 9.8 \text{ m/s}^2, x_0 = [1, -1, 0, 0]^T$$

The control objective is to track the desired trajectory given by

$$\begin{aligned} q_{1d} &= 0.4 \sin(0.4\pi t) \\ q_{2d} &= -0.5 \sin(0.5\pi t) \end{aligned}$$

The evolution of tracking errors

$$e = [e_1 \ e_2]^T = [q_1 - q_{1d} \ q_2 - q_{2d}]^T$$

In the simulation, the time horizon $[0, T]$ is partitioned into 20 subintervals. $\alpha_1 = 3$, $\alpha_2 = 1$, S and R are unit matrices of proper dimension. We first use the time-scaling transform and the control parameterization method to construct the corresponding approximated problem (Q) . Then, MISER 3.3 is utilized to solve it, giving rise to an optimal open-loop control and the corresponding optimal trajectory. Then the feedback gain matrix $K^*(t)$ is obtained by the above calculation steps.

Simulation results are shown in Figs. 2 and 3. The tracking position errors are shown in Fig. 2, and the tracking velocity errors are shown in Fig. 3.

The results given in Figs. 1 and 2 are superior to that shown in [19] using the same desired trajectory, in which an optimal neural control method of robot manipulator is proposed.

Fig. 2 Robot tracking position errors

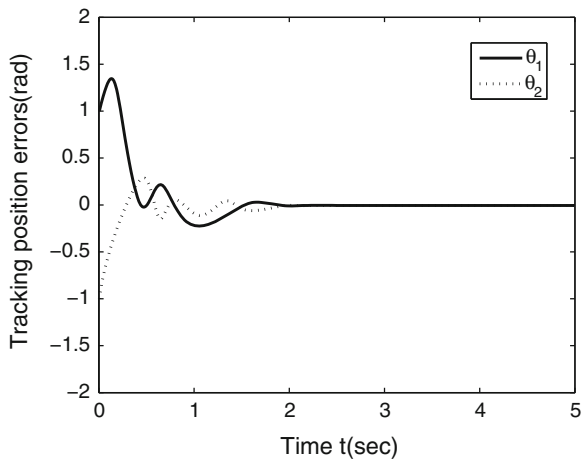
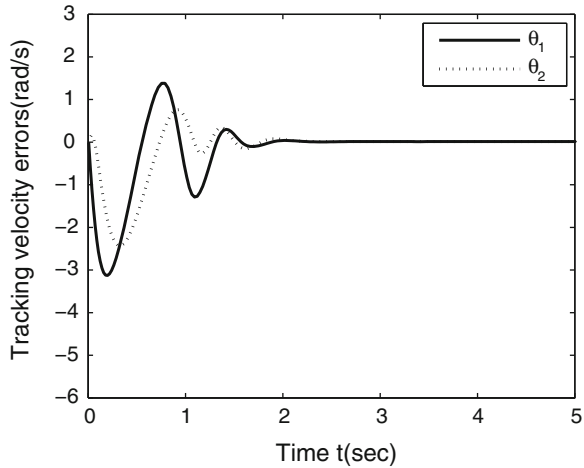


Fig. 3 Robot tracking velocity errors



5 Conclusions

The nonlinear optimal control problem of two-link RR manipulator trajectory tracking is studied in this paper, in which an optimal open-loop control is obtained firstly by using the control parametrization method and the time-scaling transform. Then the optimal closed-loop control law has been transformed into find a feedback gain matrix is required to satisfy a Riccati-like matrix differential equation, and a practical method was proposed to calculate the feedback gain matrix. The simulation results demonstrate the validity of the proposed method.

Acknowledgments This work was supported by the National Natural Science Foundation of China (61075083), and Henan Province Innovation and Technology Fund for Outstanding Scholarship (0421000500), and the Key Scientific Research Projects of Henan University of technology (09XZD008), and the Science Foundation of Henan University of Technology (150166).

References

1. Bryson AE, Meier EB (1990) Efficient algorithm for time-optimal control of a two-link manipulator. *J Guidance Control Dyn* 13:859–866
2. Constantinescu D, Croft EA (2000) Smooth and time-optimal trajectory planning for industrial manipulators along specified paths. *J Rob Syst* 17:233–249
3. Galicki M, Ucinski D (2000) Time-optimal motions of robotic manipulators. *Robotica* 18:659–667
4. Hol CWJ, Willigenburg LG, Henten EJ (2001) A new optimization algorithm for singular and non-singular digital time-optimal control of robots. In: *Proceedings of IEEE International Conference on Robotics and Automation*, pp 1136–1141

5. Park Jk, Bobrow JE (2005) Reliable computation of minimum-time motions for manipulators moving in obstacle fields using a successive search for minimum overload trajectories. *J Robot Syst* 22:1–14
6. Shiller Z (1994) On singular time-optimal control along specified paths. *IEEE Trans Robot Autom* 10:561–566
7. Banks S, McCaffrey D (1998) Lie algebras, structure of nonlinear systems and chaotic motion. *Int J Bifurcat Chaos* 8:1437–1462
8. Bobasu E, Popescu D (2006) On Modeling and multivariable adaptive control of robotic manipulators. *WSEAS Trans Syst* 5:1579–1586
9. Raimondi FM, Melluso M (2004) A neuro fuzzy controller for planar robot manipulators. *WSEAS Trans Syst* 3:2991–2996
10. Popescu D (1998) Neural control of manipulators using a supervisory algorithm, *International Conference on Automation and Quality Control, Cluj-Napoca*, pp A576–A581.
11. Huang SJ, Lee JS (2000) A stable self-organizing fuzzy controller for robotic motion control. *IEEE Trans Syst Man Cybern* 47:421–428
12. Kim YH, Lewis FL (2000) Optimal design of CMAC neural-network controller for robot manipulators. *IEEE Trans Syst Man Cybern* 30:22–28
13. Lewis FL, Yesildirek A, Liu K (1996) Multilayer neural-net robot controller with guaranteed tracking performance. *IEEE Trans Neural Networks* 7:388–396
14. Teo KL, Jennings LS, Lee HWJ, Rehbock V (1999) The control parameterization enhancing transform for constraint optimal control problems. *J Austral Math Soc Ser B* 40:314–335
15. Teo KL, Goh CJ, Wong KH (1991) *A unified computational approach to optimal control problems*. Wiley, New York
16. Zhou JY, Teo KL, Zhou D, Zhao GH (2009) Nonlinear optimal feedback control for lunar module soft landing, *IEEE International Conference on Automation and Logistics*, pp 681–688.
17. Jennings LS, Fisher ME, Teo KL, Goh CJ (1996) *MISER3.3 Optimal control software version 3.3: Theory and user manual*. Centre for Applied Dynamics and Optimization, The University of Western Australia, Australia
18. Lee HWJ, Teo KL, Yan WY (1996) Nonlinear optimal feedback control law for a class of nonlinear systems. *Parallel Sci Computations* 4:157–178
19. Nguyen TH, Pham TC (2011) Optimal neuro control of robot manipulator, *11th International Conference on Control, Automation and Systems, Gyeonggi-do, Korea*, pp 242–247

Impedance Identification and Adaptive Control of Rehabilitation Robot for Upper-Limb Passive Training

Aiguo Song, Lizheng Pan, Guozheng Xu and Huijun Li

Abstract Rehabilitation robot can assist post-stroke patients during rehabilitation therapy. The movement control of the robot plays an important role in the process of functional recovery training. Owing to the change of the arm impedance of the post-stroke patient in the passive recovery training, the conventional movement control based on PI controller is difficult to produce smooth movement to track the designed trajectory set by the rehabilitation therapist. In this paper, we model the dynamics of post-stroke patient arm as an impedance model, and an adaptive control scheme which consists of an adaptive PI control algorithm and a damp control algorithm is proposed to control the rehabilitation robot moving along predefined trajectories stably and smoothly. An equivalent 2-port circuit of the rehabilitation robot and human arm is built, and passivity theory of circuit is used to analyze the stability and smoothness performance of the robot. A slide least mean square with adaptive window (SLMS-AW) method is presented to online estimate the parameters of the arm impedance model, which is used for adjusting the gains of PI-damp controller. In this paper, the Barrett WAM Arm manipulator is used as the main hardware platform for the functional recovery training of the post-stroke patient. Passive recovery training has been implemented on the WAM Arm. Experimental results demonstrate the effectiveness and potential of the proposed adaptive control strategies.

A. Song (✉) · L. Pan · G. Xu · H. Li
School of Instrument Science and Engineering, Southeast University,
Nanjing 210096, China
e-mail: a.g.song@seu.edu.cn

L. Pan
e-mail: plz517@sina.com.cn

G. Xu
e-mail: xgzseu@yahoo.com.cn

H. Li
e-mail: lihuijun@seu.edu.cn

Keywords Rehabilitation robot · Stroke · Impedance model · Parameter identification · Robot control

1 Introduction

Stroke is a leading cause of serious, long-term disability. For instance, in China, every year there are about 2,000,000 people have a stroke, of which approximately 66 % survives the stroke, commonly involving deficits of motor function [1]. Although the optimal therapy for patients who suffer from stroke is still a point of discussion, one theory is that patients will recover better and faster when having intensive physiotherapy directly after the accident. Undamaged brain tissue will then take over the functionality of the damaged tissue and the lost functionality caused by the stroke will be regained [2]. In order to assist the stroke patients during rehabilitation therapy, some researchers have developed several robot-assisted rehabilitation therapy systems, such as MIME [3], ARM Guide [4], MIT-MANUS [5], and UECM [6]. Robotic aids can provide programmable levels of assistance and automatically modify their output based on sensor data using control frame works [7]. Rehabilitation robot usually works on two modes: one is passive recovery training mode and another is active recovery training mode. Owing to the patients exhibit a wide range of arm dysfunction levels, it is important to provide optimal assistance in robot-assisted rehabilitation, which has been demonstrated in [8]. Passive recovery training is the initial stage of rehabilitation training, and the aim of rehabilitation therapy is to reduce the muscle tone and spasticity of the impaired limb and increase its movable region [9]. The main objective in this stage is to control the robot stably and smoothly to stretch the patient to move along a predefined trajectory with the position controller. Thus, in passive recovery training mode, providing a desired movement trajectory with appropriate velocity to the patient is a key issue for robot control. Providing optimal assistance to each patient can be defined as the least amount or just enough assistance that is necessary to enable the patient to make a specific movement. Therefore, there is a need to provide controllable, quantifiable assistance specific to a particular patient by adapting the level of the assistance provided. However, there is a paucity of research in designing controllers that can realize optimal assistance. O'Malley et al. [10] used a traditional fixed gain PD trajectory controller to control the Rice Wrist to move along the desired trajectory in the GoTo mode and found that the performance was dependent on the selection of PD gains. Erol et al. [11] proposed an artificial neural network-based PI gain scheduling direct force controller which can automatically adjust control gains for a wide range of people with different conditions. Xu and Song [12] designed a fuzzy logic-based PD position controller for upper-limb rehabilitation robot to obtain stable motion tracking performance. Xu and Song [13] then developed an adaptive impedance controller based on evolutionary dynamic fuzzy neural networks for

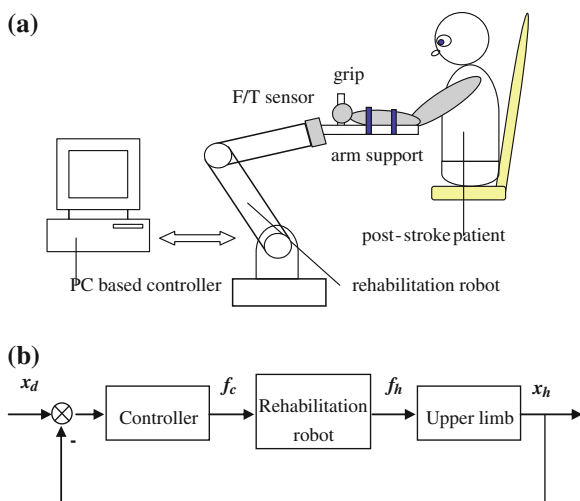
upper-limb rehabilitation robot to obtain the robust control performance when the change of impaired limb’s physical condition happens. Owing to the difficulty of neural network training and lack of sufficient training set, it is not easy to satisfy the practical need.

In this paper, a novel control scheme is proposed, which is the combination of adaptive PI control and adaptive damp control. Passivity theory is introduced to analyze the stability performance of the rehabilitation robot when interacting with patient; moreover, a SLMS-AW method is given to estimate the parameters of human arm impedance online for adjusting the gains of PI-damp controller. The Barrett WAM Arm manipulator is used as the main hardware platform for the functional recovery training of the post-stroke patient. Passive recovery training has been implemented on the WAM Arm.

2 Upper-Limb Rehabilitation Robot System

Figure 1 depicts the configuration and control structure of upper-limb rehabilitation robot system, which consists of a robot with some degrees of freedom, PC-based controller, an arm support device at the end of the robot, and a multi-dimensional force/torque sensor installed between robot end and the arm support. During the rehabilitation training, the upper limb of the post-stroke patient is banded to the arm support device at the end of the robot. The robot drives the human arm to track the designed trajectory circularly under the control of PC. The position sensor measures the movement of the robot, and the force/torque sensor measures the interactive force between the rehabilitation robot and the arm of post-stroke patient for robot control. The control structure of position tracking can be shown as Fig. 1b.

Fig. 1 Upper-limb rehabilitation robot and control structure.
a Rehabilitation system.
b Control structure



3 Modeling of Rehabilitation Robot System

3.1 Dynamics of the Rehabilitation Robot

For the simplification of the analysis, we only consider the one DOF case of the rehabilitation robot; the results of the analysis are easy to expand to the n-DOF cases.

The robot dynamics can be expressed as follows:

$$f_c(t) - f_h(t) = m_r \ddot{x}_r(t) + b_r \dot{x}_r(t) + k_r x_r(t) \quad (1)$$

where f_c is output of controller as a force command, x_r is displacement of the robot, and m_r , b_r , and k_r stand for mass, damp, and stiffness of the robot, respectively. f_h is applied force on the human arm by robot.

Equation (1) can be expressed in frequency domain by using Laplace transformation as

$$f_c(s) - f_h(s) = (m_r s + b_r + k_r/s) \dot{x}_r(s) \quad (2)$$

Let

$$Z_r(s) = m_r s + b_r + k_r/s \quad (3)$$

$Z_r(s)$ is mechanical impedance of the robot.

3.2 Impedance Model of the Human Arm

The human upper-limb dynamics is usually expressed by a mass-spring-damp system as follows:

$$f_h(t) - f_a(t) = m_h \ddot{x}_h(t) + b_h \dot{x}_h(t) + k_h x_h(t) \quad (4)$$

where f_a is the active force of the post-stroke patient. During the passive recovery training mode, f_a is sometimes caused by spastic muscle, which can be treated as an interference force affecting on the rehabilitation system. x_h is displacement of the human arm, and m_h , b_h , and k_h stand for mass, damp, and stiffness of the human arm, respectively.

Rewriting Eq. (4) in frequency domain as

$$f_h(s) = Z_h(s) \dot{x}_h(s) + f_a(s) \quad (5)$$

$Z_h(s) = m_h s + b_h + k_h/s$ is mechanical impedance of the human arm. During the rehabilitation training process, owing to the changes of wrist joint, elbow joint, and shoulder joint when human upper limb is passively driven by the robot, the m_h , b_h , k_h are changed continuously and circularly. Thus, the mechanical impedance of

the human upper limb is typically time-varying. Because it is impossible to pre-known the change of the $Z_h(s)$ and when the f_a is happen accurately, the environment of rehabilitation robot is parameter uncertain.

3.3 Control of Rehabilitation Robot

Conventional PI controller is used for position trajectory control. In order to keep the trajectory of human arm smooth and stable without abrupt change in velocity, damp control is incorporated with PI control, which can prevent robot from high speed and thus maintains the safety of the rehabilitation robot system. In this paper, the block diagram of suggested adaptive PI-damp controller is shown in Fig. 2.

The outputs of PI control and damp control are given as following:

$$f_{PI} = K_P(\dot{x}_d - \dot{x}_h) + K_I(x_d - x_h) \tag{6}$$

$$f_{damp} = -R_d\dot{x}_h \tag{7}$$

$$f_c = f_{PI} + f_{damp} = K_P(\dot{x}_d - \dot{x}_h) + K_I(x_d - x_h) - R_d\dot{x}_h \tag{8}$$

Here, K_P and K_I stand for proportional coefficient and integral coefficient of PI controller, respectively. R_d stands for damp coefficient which is a parameter of damp controller.

3.4 Equivalent 2-Port Circuit of Rehabilitation Robot

According to the equivalent rule between mechanical system and electrical system, such as current I is equivalent to velocity \dot{x} , and voltage U is equivalent to force F , we can express Eq. (8) in frequency domain as

$$U_c = U_{PI} + U_{damp} = Z_c(I_d - I_h) - R_d I_h \tag{9}$$

where $Z_c(s) = K_p + K_I/s$ is defined as control impedance of the robot. Therefore, the rehabilitation robot can be depicted as an equivalent 2-port circuit based on Eqs. (1)–(9), seen in Fig. 3.

Fig. 2 Adaptive PI and damp controllers

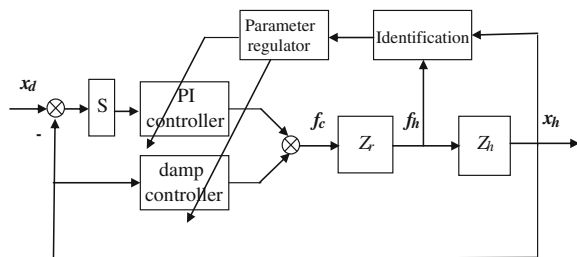
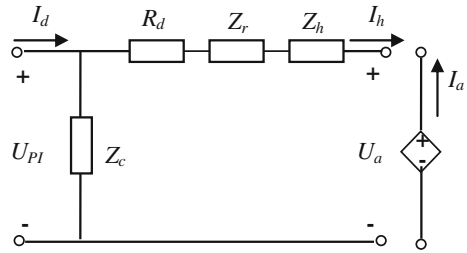


Fig. 3 Equivalent 2-port circuit of rehabilitation robot



where, for the convenience of analysis, the interference force U_a caused by patient is treated as an input voltage source of the 2-port circuit, and $I_a(s) = -I_h(s)$ stands for the input velocity.

4 Analysis of Stability and Smoothness Performance

4.1 Passivity Performance Analysis

Definition 1 [14] An n-port circuit is said to be passive if and only if for any independent set of n-port flows, I_i injected into the circuit, and efforts, U_i applied across the circuit

$$\int_0^\infty U^T(t)I(t)dt \geq 0 \tag{10}$$

where $U^T = [U_1, U_2, \dots, U_n]^T \in L_2^n(\mathbb{R}^+)$ and $I^T = [I_1, I_2, \dots, I_n]^T \in L_2^n(\mathbb{R}^+)$.

Condition (10) is simply a statement that a passive n-port circuit may dissipate energy but cannot increase the total energy of a system in which it is an element. The passivity of the circuit implies the stability of the system.

Assumption The parameters of robot impedance $m_r, b_r, k_r \in \mathbb{R}^+$, are fixed, and the parameters of human arm impedance $m_h, b_h, k_h \in \mathbb{R}^+$ are bounded $m_h \leq \lambda_m, b_h \leq \lambda_b, k_h \leq \lambda_k$. The parameters of controller $K_P, K_I, R_d \in \mathbb{R}^+$ are bounded.

Let

$$Z_d(s) = R_d + Z_r + Z_h = (m_r + m_h)s + (R_d + b_r + b_h) + \frac{1}{s}(k_r + k_h) \tag{11}$$

Thus, $Z_d(s)$ is a typical energy dissipation impedance.

The relationship between effort $U(t)$ (force, voltage) and flow $I(t)$ (velocity, current) of the equivalent 2-port circuit of rehabilitation robot can be conveniently specified by its hybrid matrix $H(s)$ according to

$$\begin{bmatrix} U_{PI} \\ I_a \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} I_d \\ U_a \end{bmatrix} = H(s) \begin{bmatrix} I_d \\ U_a \end{bmatrix} \quad (12)$$

Deducing from the Eqs. (1)–(9) and (11), we have

$$\begin{bmatrix} U_{PI} \\ I_a \end{bmatrix} = \begin{bmatrix} \frac{Z_c Z_d}{Z_c + Z_d} & \frac{Z_c}{Z_c + Z_d} \\ -\frac{Z_c}{Z_c + Z_d} & \frac{1}{Z_c + Z_d} \end{bmatrix} \begin{bmatrix} I_d \\ U_a \end{bmatrix} \quad (13)$$

$$H(s) = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} = \begin{bmatrix} \frac{Z_c Z_d}{Z_c + Z_d} & \frac{Z_c}{Z_c + Z_d} \\ -\frac{Z_c}{Z_c + Z_d} & \frac{1}{Z_c + Z_d} \end{bmatrix} \quad (14)$$

So, for the equivalent 2-port circuit of rehabilitation robot, we have

$$\begin{aligned} \int_0^{\infty} [U_{PI} \quad I_a] H^T [I_d \quad U_a]^T dt &= \int_0^{\infty} (h_{11} I_d^2 + (h_{12} + h_{21}) I_d U_a + h_{22} U_a^2) dt \\ &= \int_0^{\infty} \left(\frac{Z_c Z_d}{Z_c + Z_d} I_d^2 + \frac{1}{Z_c + Z_d} U_a^2 \right) dt \geq 0 \end{aligned} \quad (15)$$

Therefore, the rehabilitation robot system under PI-damp control with bounded parameters is always passive, which means it is stable.

4.2 Smooth Position Tracking Performance Analysis

Assuming the interference voltage $U_a = 0$, the movement of the human arm

$$I_h = \frac{Z_c}{Z_c + Z_d} = \frac{Z_c}{\tilde{Z}} I_d = \frac{(K_P s + K_I)}{\tilde{m} s^2 + \tilde{b} s + \tilde{k}} I_d \quad (16)$$

Let

$$\begin{aligned} \tilde{Z} = Z_c + Z_d &= (m_r + m_h) s + (R_d + b_r + b_h + K_P) + \frac{1}{s} (k_r + k_h + K_I) \\ &= \tilde{m} s + \tilde{b} + \tilde{k}/s \end{aligned} \quad (17)$$

where

$$\tilde{m} = m_r + m_h; \quad \tilde{b} = R_d + b_r + b_h + K_P; \quad \tilde{k} = k_r + k_h + K_I \quad (18)$$

The position tracking error between the desire trajectory and real trajectory of robot is

$$e = I_d - I_h = I_d - \frac{(K_P s + K_I)}{\tilde{m} s^2 + \tilde{b} s + \tilde{k}} I_d = \frac{\tilde{m} s^2 + (\tilde{b} - K_P) s + (\tilde{k} - K_I)}{\tilde{m} s^2 + \tilde{b} s + \tilde{k}} I_d \quad (19)$$

Thus, the steady-state position tracking error

$$e_{ss} = \lim_{s \rightarrow 0} se(s) \frac{1}{s} = \frac{\tilde{k} - K_I}{\tilde{k}} = \frac{k_h}{K_I + k_h} \quad (20)$$

For the interference $f_a(t) = \delta(t)$ caused by patient, the position tracking error

$$e' = \frac{1}{\tilde{m}s^2 + \tilde{b}s + \tilde{k}} U_a \quad (21)$$

So, the steady-state position tracking error caused by interference U_a

$$e'_{ss} = \lim_{s \rightarrow 0} se'(s) = 0 \quad (22)$$

The above equation means the control structure is insensitive to the interference U_a .

Let $\omega_n = \sqrt{\frac{\tilde{k}}{\tilde{m}}}$, $\zeta = \frac{1}{2} \frac{\tilde{b}}{\sqrt{\tilde{m}\tilde{k}}}$, and substitute them into Eq. (16),

$$G(s) = \frac{I_h}{I_d} = \frac{\frac{\omega_n^2}{k} (K_P s + K_I)}{s^2 + 2\zeta\omega_n s + \omega_n^2} \quad (23)$$

From the theory of automatic control, if $\zeta \geq 1$, the control system is called damping system.

$$\tilde{b} \geq 2\sqrt{\tilde{m}\tilde{k}} \quad (24)$$

In this case, there is no overshoot in the step response of the system, which means the smoothness performance is desired for post-stroke patient passive recovery training. So that the parameters of the controller should satisfy the smoothness condition as

$$R_d + b_r + b_h + K_P \geq 2\sqrt{(m_r + m_h)(k_r + k_h + K_I)} \quad (25)$$

5 Identification of Impedance Model

The impaired limb's dynamics can be expressed as a time-variant mass-spring-damper model. Suppose \hat{m}_h , \hat{b}_h , \hat{k}_h are estimates of m_h , b_h , k_h in Eq. (4), respectively, we have

$$\hat{f}_h = \hat{m}_h \ddot{x}_h + \hat{b}_h \dot{x}_h + \hat{k}_h x_h \quad (26)$$

According to the least mean square method,

$$E = \sum_{i=1}^N [f_h(i) - \hat{f}_h(i)]^2 \quad (27)$$

$$\frac{\partial E}{\partial \hat{m}_h} = 0; \quad \frac{\partial E}{\partial \hat{b}_h} = 0; \quad \frac{\partial E}{\partial \hat{k}_h} = 0 \tag{28}$$

So, we have

$$\begin{bmatrix} \hat{m}_h \\ \hat{b}_h \\ \hat{k}_h \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^N \ddot{x}_h^2(i) & \sum_{i=1}^N \ddot{x}_h(i)\dot{x}_h(i) & \sum_{i=1}^N \dot{x}_h(i)x_h(i) \\ \sum_{i=1}^N \dot{x}_h(i)\ddot{x}_h(i) & \sum_{i=1}^N \dot{x}_h^2(i) & \sum_{i=1}^N \dot{x}_h(i)x_h(i) \\ \sum_{i=1}^N x_h(i)\ddot{x}_h(i) & \sum_{i=1}^N x_h(i)\dot{x}_h(i) & \sum_{i=1}^N x_h^2(i) \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^N \ddot{x}_h(i)f_h(i) \\ \sum_{i=1}^N \dot{x}_h(i)f_h(i) \\ \sum_{i=1}^N x_h(i)f_h(i) \end{bmatrix} \tag{29}$$

N is number of sampling points for parameter estimation. In order to estimate the parameters online, we have proposed a kind of slide least mean square (SLMS) method [15] as follows:

$$\begin{bmatrix} \hat{m}_h(t) \\ \hat{b}_h(t) \\ \hat{k}_h(t) \end{bmatrix} = \begin{bmatrix} \sum_{i=t-N+1}^t \ddot{x}_h^2(i) & \sum_{i=t-N+1}^t \ddot{x}_h(i)\dot{x}_h(i) & \sum_{i=t-N+1}^t \dot{x}_h(i)x_h(i) \\ \sum_{i=t-N+1}^t \dot{x}_h(i)\ddot{x}_h(i) & \sum_{i=t-N+1}^t \dot{x}_h^2(i) & \sum_{i=t-N+1}^t \dot{x}_h(i)x_h(i) \\ \sum_{i=t-N+1}^t x_h(i)\ddot{x}_h(i) & \sum_{i=t-N+1}^t x_h(i)\dot{x}_h(i) & \sum_{i=t-N+1}^t x_h^2(i) \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=t-N+1}^t \ddot{x}_h(i)f_h(i) \\ \sum_{i=t-N+1}^t \dot{x}_h(i)f_h(i) \\ \sum_{i=t-N+1}^t x_h(i)f_h(i) \end{bmatrix} \tag{30}$$

$$[\hat{Z}_h(t)] = [A(t)]^{-1}[C(t)] \quad t \geq N \tag{31}$$

The elements of the matrixes $[A(t)]$ and $[C(t)]$ can be quickly calculated by using slide method as

$$\begin{aligned} a_{i,j}(t+1) &= \sum_{k=t-N+2}^{t+1} x_h^{(3-i)}(k)x_h^{(3-j)}(k) \\ &= a_{i,j}(t) + x_h^{(3-i)}(t+1)x_h^{(3-j)}(t+1) - x_h^{(3-i)}(t-N+1)x_h^{(3-j)}(t-N+1) \\ & \quad i = 1, 2, 3; j = 1, 2, 3 \end{aligned} \tag{32}$$

$$\begin{aligned} c_i(t+1) &= \sum_{k=t-N+2}^{t+1} x_h^{(3-i)}(k)f_h(k) \\ &= c_i(t) + x_h^{(3-i)}(t+1)f_h(t+1) - x_h^{(3-i)}(t-N+1)f_h(t-N+1) \quad i = 1, 2, 3 \end{aligned} \tag{33}$$

In general, the parameter N is fixed. In this research, in order to estimate the impaired limb's parameters more effectively and real-time, a SLMS with adaptive window (SLMS-AW) identification algorithm is given. N is dynamically adapted according to the variations of f_h and \ddot{x}_h , which can be expressed as

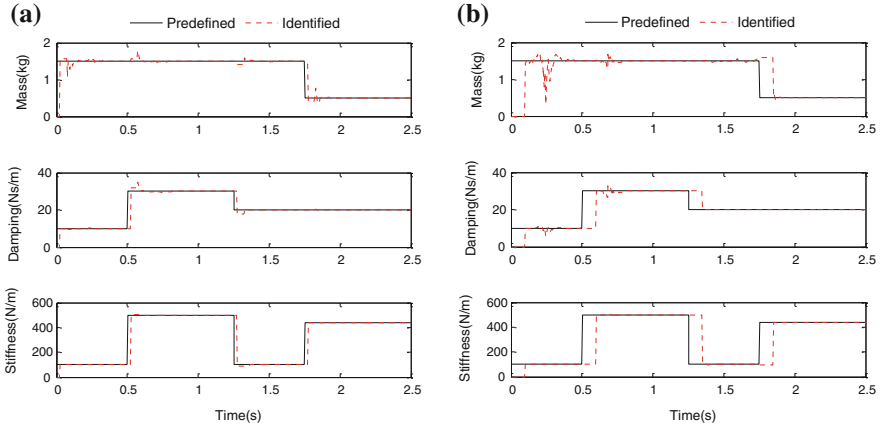


Fig. 4 Identification results of human arm impedance. **a** Identification using SLMS-AW. **b** Identification using traditional SLMS

$$\begin{aligned}
 N &= f(\Delta\ddot{x}_h, \Delta f_h) \quad N \in [N_{\min}, N_{\max}] \\
 f(\Delta\ddot{x}_h, \Delta f_h) &= N_{\max} - (N_{\max} - N_{\min}) \cdot \left[(1 - \lambda) \frac{\Delta\ddot{x}_h}{\Delta\ddot{x}_{\max}} + \lambda \frac{\Delta f_h}{\Delta f_{\max}} \right] \quad (34)
 \end{aligned}$$

where N_{\min}, N_{\max} are the up and low limitations of N , respectively; $\Delta\ddot{x}_{\max}$ and Δf_{\max} are the maximum variations of accelerator and force; and λ is the weight coefficient.

To verify the performance of suggested SLMS-AW, some simulation experiments have been carried out. The results of identification of mass, damp, and stiffness of human arm impedance are shown in Fig. 4. For comparison, the identification results of traditional SLMS identification algorithm are also given. In Fig. 4, solid line represents the change curve of parameters of human arm impedance which are predefined according to the experimental results and clinical analysis in [16, 17], and dashed line represents the identification curve. From the identification results in Fig. 4, it can be concluded that the proposed SLMS-AW algorithm is obviously more accurate, robust and real-time than traditional SLMS method.

6 Experiment

6.1 Experimental Setup

A 4-DOF Barrett WAM Arm manipulator shown in Fig. 5 is used as the main hardware platform for the functional recovery therapy in this research. The upper-limb rehabilitation experimental setup consists of the Barrett WAM Arm, a three-dimensional force sensor (Fig. 5a), an arm support device (Fig. 5b), and an

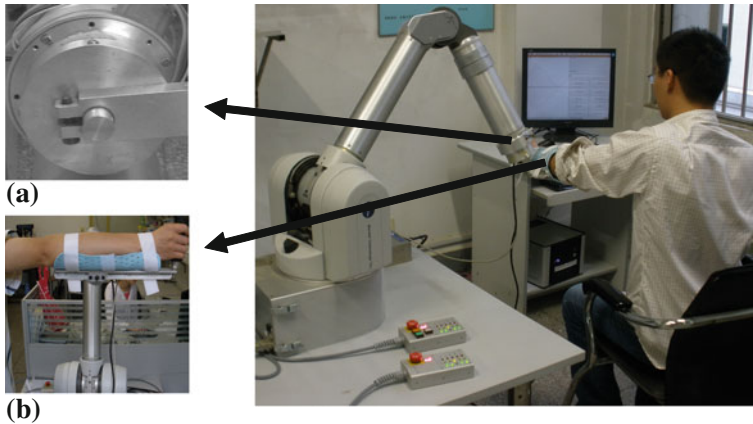


Fig. 5 Rehabilitation robot systems based on WAM arm. **a** Three-dimensional sensor. **b** Arm support device

external PC offered by Barrett. In order to record the force between the patient arm and the rehabilitation robot end-effector, a three-dimensional force sensor [18] is designed and installed at the end-effector of the WAM. With the arm support device, the forearm of the patient can be well supported on it. An external PC running with the Linux system was responsible for running the control loop at 2 kHz and providing high-level command of the WAM rehabilitation system. Following the generated high-level command, the upper limb of the patient could be stretched by the WAM rehabilitation robot and could perform various physical trainings. All real-time communication between the external PC and the motor Pucks is done via an internal high-speed CAN bus.

6.2 Control System

During the robot-assisted passive recovery training, the physical state of patient’s upper limb is not ideal; there are many uncertain factors affecting the control performance, for example, pose-position change, muscle spasm and tremor, even occasional cough, and other external disturbance. In this research, the adaptive PI control algorithm and damp control algorithm are proposed for the passive recovery training to control the WAM Arm stably and smoothly to stretch the impaired limb to move along the predefined trajectory. The adaptive PI-damp controller is expected to provide better performance than traditional fixed gain PI controller because of its ability of adjusting the control gains in accordance with changes of the patient’s physical state. The block diagram of proposed control strategy is given in Fig. 6.

As shown in Fig. 6, the designed control system mainly consists of controller and identification unit and parameter regulators. The designed controller adopts

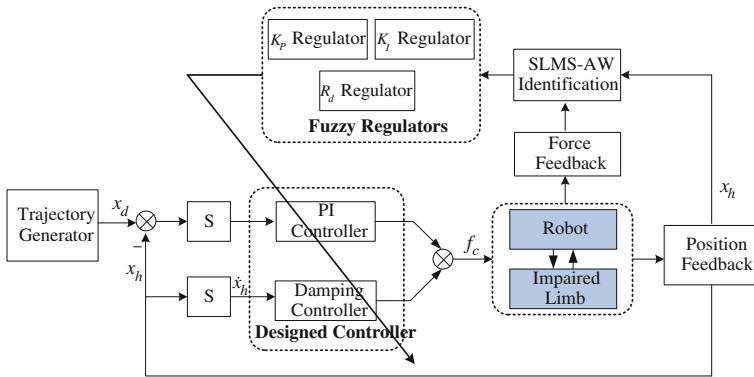


Fig. 6 Control system block diagram for passive training

adaptive PI-damp control algorithms. In this research, damp control works as an energy dissipation part, which is proportional to velocity with opposite direction and is incorporated with PI control. Under the designed controller, the rehabilitation robot stretches the impaired limb to do recovery training. The proposed control method can effectively prevent robot from high speed and interference caused by post-stroke patient, so that the robot can run with the stable and smooth movement tracking during rehabilitation training. The SLMS-AW identification unit estimates the impaired-limb impedance parameters online, which stand for the dynamic state of the training arm. Then, the parameter regulators adapt the K_P , K_I , and R_d of designed controller according to the arm physical condition.

In this paper, fuzzy reasoning logic is adopted to adjust the controller parameters. There are three separate fuzzy regulators for K_P , K_I , and R_d parts, respectively. During the rehabilitation exercise, the \hat{m}_h , \hat{b}_h , \hat{k}_h of training limb are estimated by SLMS-AW identification. In order to regulate the control parameters effectively and appropriately, the identified results of mass and stiffness are used for adjusting K_P , and identified results of mass and damping are used for adjusting K_I . Considering the function of damp control, the identified damping and measured active force of subject are selected to regulate R_d . All the inputs of fuzzy regulators are scaled to $[0,1]$, and the corresponding outputs are separately scaled: $K_P \in [500, 700]$, $K_I \in [650, 900]$, and $R_d \in [0, 50]$. Meanwhile, during the fuzzification and defuzzification, all the inputs and outputs are defined as five fuzzy sets: small (S), small and middle (SM), middle (M), middle and large (ML), and large (L). According to the practical application, the K_P regulator should adjust gently and be less affected to mass change, while the R_d regulator should be magnificently sensitive to the active force. The designed fuzzy reasoning rules are shown in Table 1, 2, and 3, respectively, for regulating K_P , K_I , and R_d , and the corresponding input–output surface maps are shown in Fig. 7, 8, and 9.

Table 1 Fuzzy reasoning rules for K_P

m-scaled	k-scaled				
	S	SM	M	ML	L
S	S	S	S	SM	M
SM	S	SM	M	M	ML
M	S	SM	M	M	ML
ML	SM	SM	M	ML	ML
L	M	M	ML	ML	L

Table 2 Fuzzy reasoning rules for K_I

m-scaled	b-scaled				
	S	SM	M	ML	L
S	S	S	SM	SM	M
SM	S	SM	SM	M	M
M	SM	SM	SM	M	ML
ML	SM	M	M	ML	ML
L	M	M	ML	ML	L

Table 3 Fuzzy reasoning rules for R_d

f-scaled	b-scaled				
	S	SM	M	ML	L
S	S	S	S	SM	SM
SM	SM	SM	SM	SM	M
M	M	M	M	M	ML
ML	ML	M	ML	ML	L
L	ML	L	L	L	L

Fig. 7 Control surface of K_P

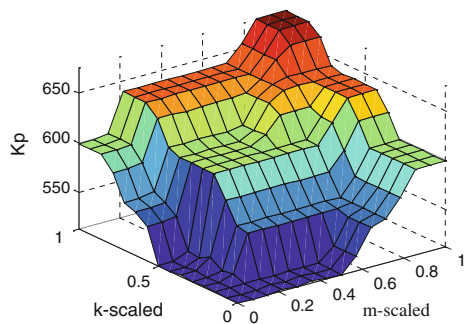


Fig. 8 Control surface of K_I

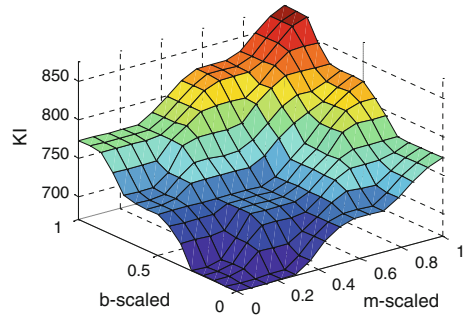
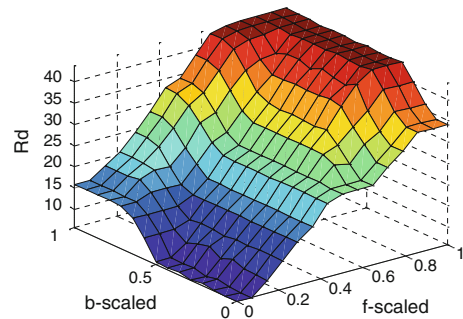


Fig. 9 Control surface of R_d



6.3 Experimental Results

To verify the effectiveness of the proposed adaptive PI-damp controller, two healthy subjects are recruited to participate in the robot-aided upper-limb passive rehabilitation exercise. Figure 10 shows the sinusoidal trajectory tracking performances of proposed adaptive PI-damp strategy and traditional PI method with regard to a healthy subject sample P1 in the horizontal (J3, the 3rd joint) flexion/extension exercises. The human arm's flexion/extension limitations expressed in WAM world frames are defined as -0.8 rad in flexion and 0.8 rad in extension, respectively. Conventional PI controller gains K_P , K_I , and R_d were set as 600, 750, and 1.5, respectively. The safety peak joint velocity and the maximum motor torque for four motors were set as 1.2 rad/s and 8.2 Nm, respectively. The inspection of Fig. 10 shows that when the participant is guided by the WAM along the predefined sinusoidal trajectory, both the adaptive PI-damp controller and traditional PI controller can achieve the desired trajectories, but the tracking performance of the former is even better than the one for the latter which can be observed from trajectory tracking error.

Meanwhile, Maximum absolute error (MAE) and sum of absolute error (SAE) of trajectory tracking are selected as two indices to quantitatively evaluate the performances of the new and conventional methods. Table 4 gives the quantitative

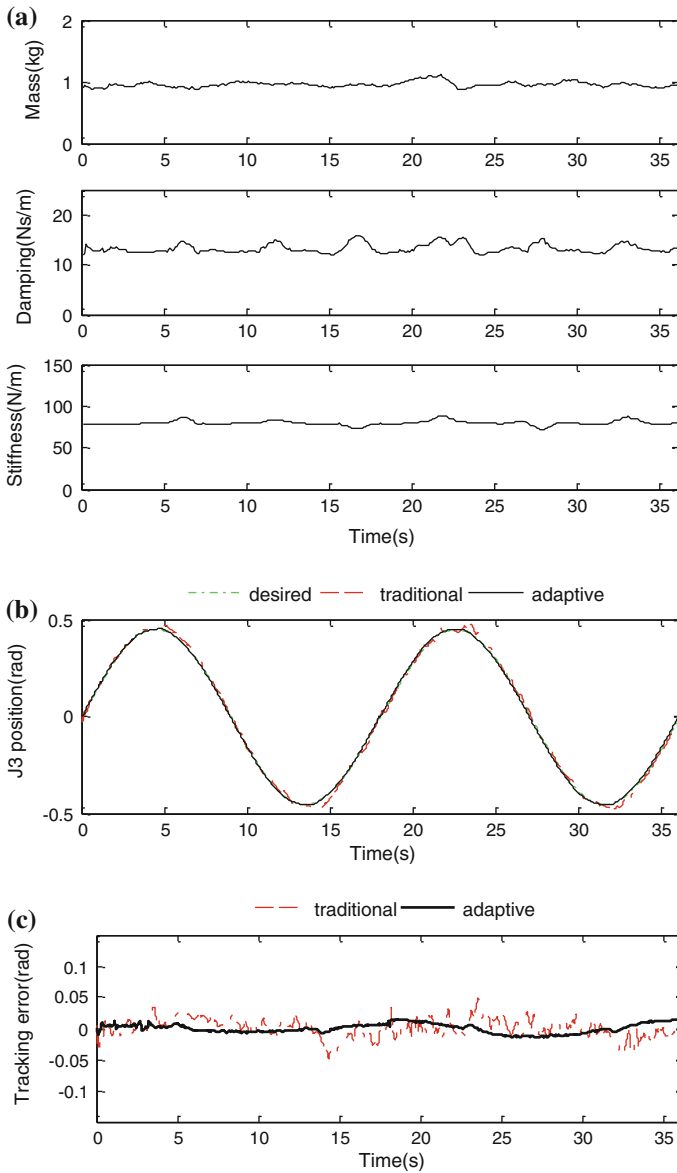


Fig. 10 Results of trajectory track control for P1 in stationary state. **a** Arm identification in stationary. **b** Trajectory tracking in stationary. **c** Tracking error in stationary

comparison results with two indices. It is shown that the MAE (0.017865) and SAE (28.446) of the new methods are obviously smaller than the ones (0.049066, 44.022) for the conventional method. A further comparison is made under the non-stationary condition that the participant intentionally applies disturbance force at

Table 4 Control performance comparison for P1

Tracking error (rad)		MAE	SAE
Stationary	Adaptive	0.017865	28.446
	Traditional	0.049066	44.022
Non-stationary	Adaptive	0.059341	37.835
	Traditional	0.10919	52.993

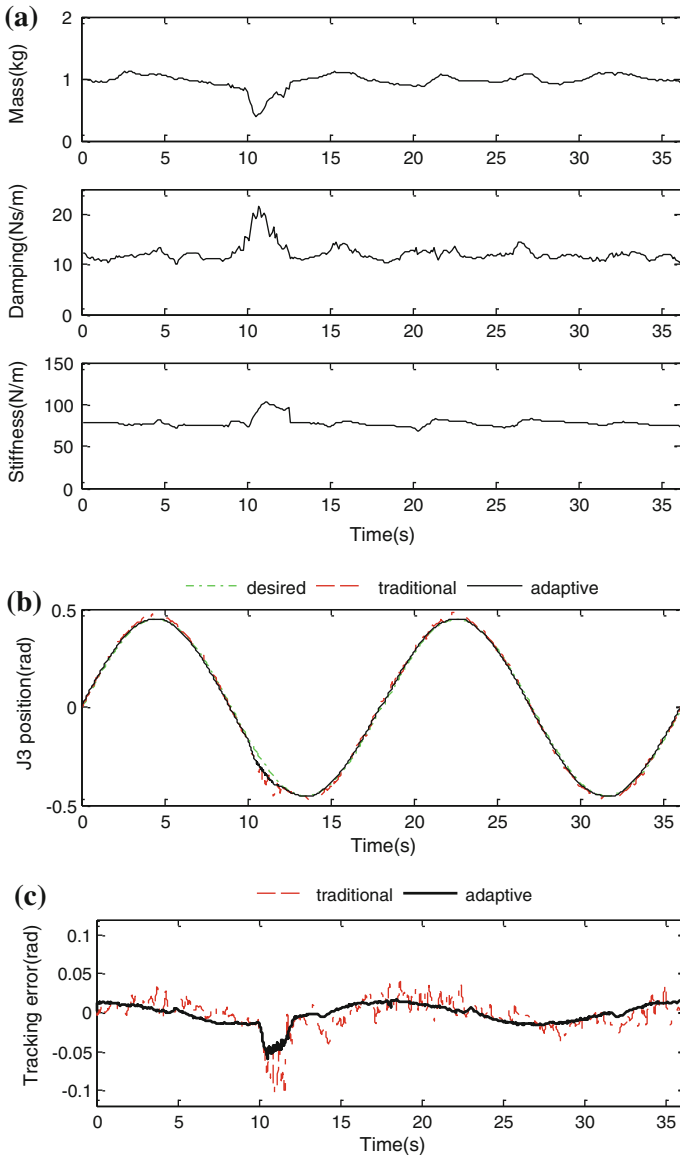


Fig. 11 Results of trajectory track control for P1 in non-stationary state. **a** Arm identification in non-stationary. **b** Trajectory tracking in non-stationary. **c** Tracking error in non-stationary

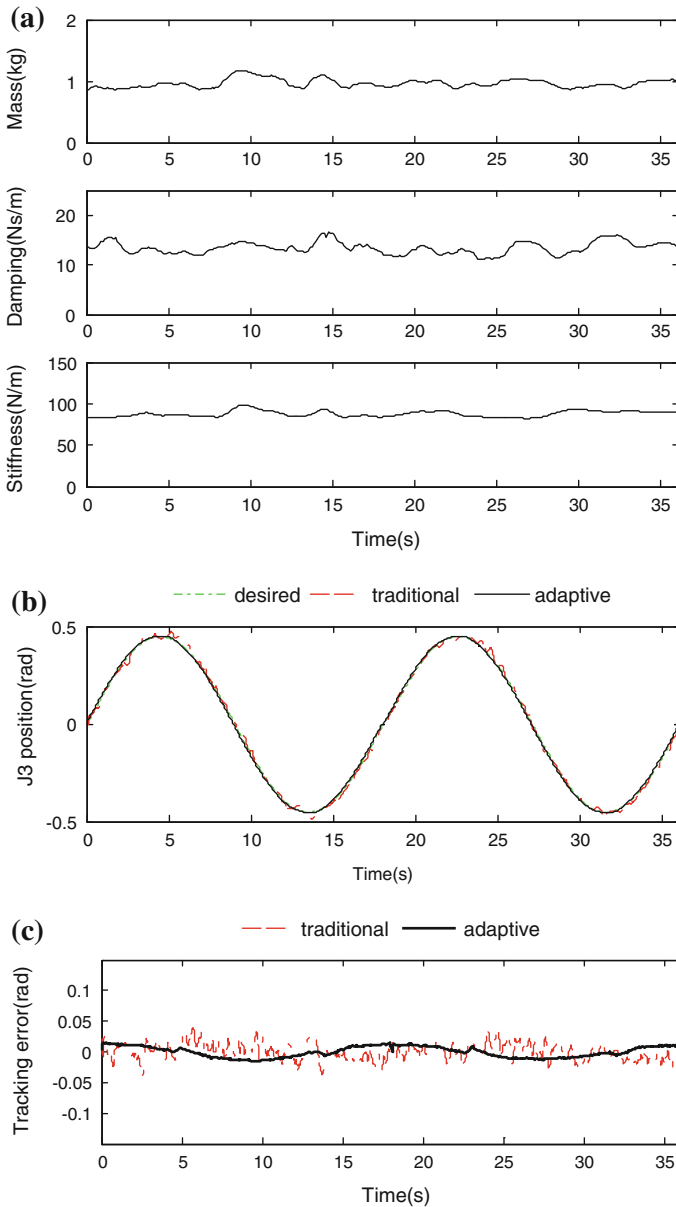


Fig. 12 Results of trajectory track control for P2 in stationary state. **a** Arm identification in stationary. **b** Trajectory tracking in stationary. **c** Tracking error in stationary

the time interval from 10 to 12 s. Figure 11 shows the representative results. From the human arm's impedance identification profile, when the subject is asked to apply intentional force, the estimated human arm's mass, damping, and stiffness

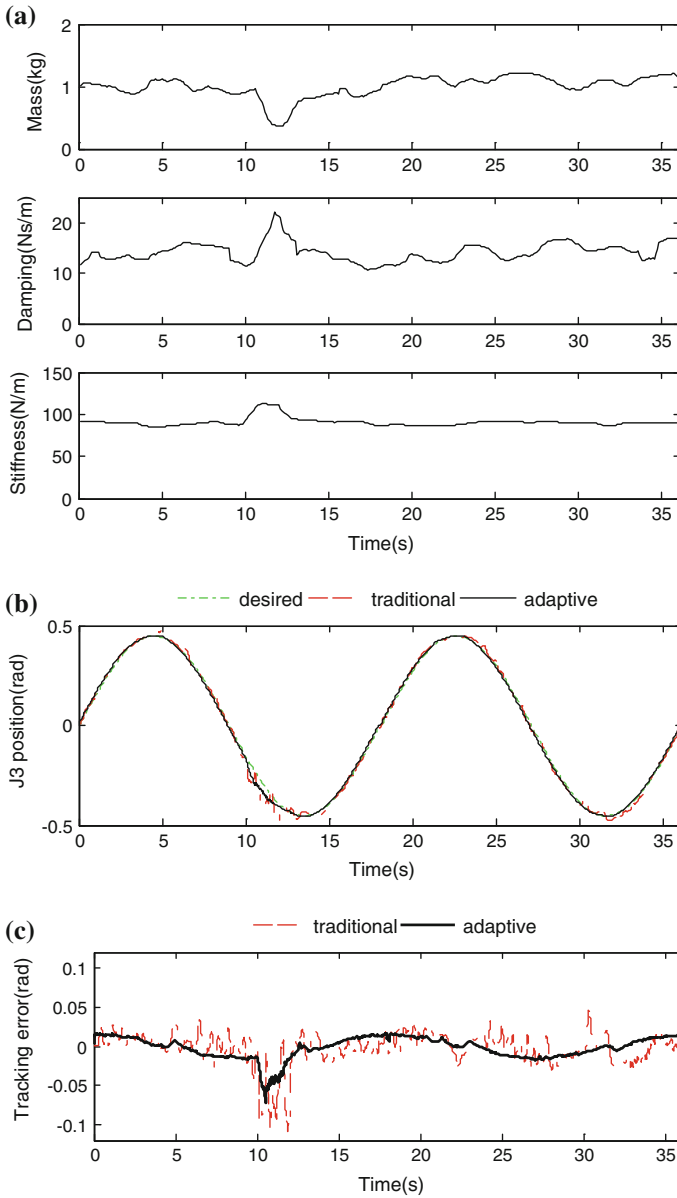


Fig. 13 Results of trajectory track control for P2 in non-stationary state. **a** Arm identification in non-stationary. **b** Trajectory tracking in non-stationary. **c** Tracking error in non-stationary

parameters show substantial increases at the moment when the arm muscle force increases intentionally. Although a certain interference forces are exerted the passive rehabilitation training system, the sinusoidal trajectory tracking of the

Table 5 Control performance comparison for P2

Tracking error (rad)		MAE	SAE
Stationary	Adaptive	0.015636	23.261
	Traditional	0.045019	39.651
Non-stationary	Adaptive	0.072641	38.507
	Traditional	0.10909	51.475

adaptive control method is still well achieved. Moreover, it is observed that the joint position control smoothness of adaptive PI-damp strategy is obviously superior to the one from the traditional PI control method which can also be found from the MAE and SAE shown in Table 4.

To verify the adaptability of the proposed control method among different subjects, another participant P2 is asked to perform the same passive rehabilitation exercises as the one conducted for participant P1. Figures 12 and 13 show the P2's arm impedance parameter identification and trajectory tracking performance. Corresponding quantitative results for MAE and SAE are also illustrated in Table 5. The quantitative analysis of trajectory tracking errors of two subjects in Tables 4 and 5 shows that the errors of proposed control strategy are obviously less than that of the traditional PI control method. Therefore, the experimental results demonstrate that the proposed adaptive PI-damp controller has better performance of stability and smoothness than the conventional controller.

7 Conclusions

In this paper, an adaptive PI-damp controller is proposed for rehabilitation robot control, considering that the training-limb physical condition is usually dynamic during the passive recovery exercise. The models of rehabilitation robot and human upper limb are built, and then, the stability and smoothness performance of the robot is analyzed by using passivity theory. The SLMS-AW method is given to identify the parameters of training limb in real time, which represent the upper-limb physical condition. Three fuzzy regulators are designed to adapt the parameters of PI-damp controller according to the real-time physical state of training limb. Experimental results obtained on the Barrett WAM Arm hardware platform demonstrate that the proposed adaptive PI-damp control strategy has better performance of stability and smoothness than the traditional PI algorithm.

Acknowledgments This work was supported by the National Natural Science Foundation of China (No. 61272379, 61104206), the Natural Science Foundation of JiangSu Province (BK2010063), and Foundation of ChangZhou (CE20120085).

References

1. Homepage of Chinese Ministry of Health www.moh.gov.cn/public/
2. Michel VEG, Driessen BJF, Michel D et al (2005) A Motorized gravity compensation mechanism used for Active Rehabilitation of upper limbs. In: Proceedings of the 2005 IEEE 9th international conference on rehabilitation robotics, Chicago, pp 152–155
3. Burgar CG, Lum PS, Shor PC et al (2000) Development of robots for rehabilitation therapy: the Polo Altova/Stanford experience. *J Rehabil Res Dev* 37(6):663–673
4. Reinkensmeyer DJ, Kahn LE, Arerbuch M et al (2000) Understanding and treating arm movement impairment after chronic brain injury: Progress with the ARM Guide. *J Rehabil Res Dev* 37(6):653–662
5. Krebs HI, Volpe BT, Aisen ML, Hogan N (2000) Increasing productivity and quality of care: robot-aided neuro-rehabilitation. *J Rehabil Res Dev* 37(6):639–652
6. Zhang YB, Wang ZX, Ji LH (2005) The clinical application of the upper extremity compound movements rehabilitation training robot. In: Proceedings of the 2005 IEEE 9th international conference on rehabilitation robotics, Chicago, pp 91–94
7. Krebs HI, Hogan N, Aisen ML, Volpe BT (1998) Robot-aided neurorehabilitation. *IEEE Trans On Rehab Eng* 6:75–87
8. Kahn LE, Rymer WZ, Reinkensmeyer DJ (2004) Adaptive assistance for guided force training in chronic stroke. In: Proceedings of the 26th annual international conference of the IEEE EMBS, San Francisco, pp 2722–2725
9. Lindberg P, Schmitz C, Forssberg H (2004) Engardt: Effects of passive-active movement training on upper limb motor function and cortical activation in chronic patients with stroke: a pilot study. *J Rehabil Med* 36:117–123
10. O'Malley MK, Sledd A, Gupta A (2006) The rice wrist: a distal upper extremity rehabilitation robot for stroke therapy. In: Proceedings IMECE, Chicago, pp 1–10
11. Duygun E, Vishnu M, Nilanjan S et al (2005) A new control approach to robot assisted rehabilitation. In: Proceedings of the 2005 IEEE 9th international conference on rehabilitation robotics, Chicago, pp 323–328
12. Xu GZ, Song AG, Li HJ (2011) Control system design for an upper-limb rehabilitation robot. *Adv Rob* 25(1):229–251
13. Xu GZ, Song AG, Li HJ (2011) Adaptive impedance control for upper-limb rehabilitation robot using evolutionary dynamic recurrent fuzzy neural networks. *J Intell Rob Syst* 62(2):501–525
14. Anderson RJ, Spong MW (1989) Bilateral control of teleoperators with time delay. *IEEE Trans Autom Control* 34(5):494–501
15. Li HJ, Song AG (2007) Virtual-environment modeling and correction for force-reflecting teleoperation with time delay. *IEEE Trans Ind Elec* 54(2):1227–1233
16. Lin CC, Ju MS, Lin CW et al (2003) The pendulum test for evaluating spasticity of the elbow joint. *Arch Phys Med Rehabil* 84:69–74
17. Noritsugu T, Tanaka T (1997) Application of rubber artificial muscle manipulator as a rehabilitation robot. *IEEE/ASME Trans Mechatronics* 2(4):259–267
18. Song AG, Wu J, Qin G, Huang WY (2007) A novel self-decoupled four degree-of-freedom wrist force/torque sensor. *Measurement* 40(9): 883–889

Independent Component Analysis: Embedded LTSA

Qianwen Yang, Yuan Li, Fuchun Sun and Qingwen Yang

Abstract In order to solve the adaptability problem of local tangent space alignment (LTSA) with potential higher-order information loss in manifolds, a novel algorithm is proposed to optimize the extraction of local neighborhood information. The algorithm is based on LTSA and ICA algorithms, which is called the IELTSA algorithm. By optimizing the extraction of the tangent vectors, the algorithm can improve dimension reduction in high-dimensional and unevenly distributed manifolds. The proposed algorithm is feasible in carrying out manifold learning, and the reconstruction error is no more than that of LTSA. Experiments show that IELTSA can be applied to changed-density 3D curves and face images targeted at lower dimensions, showing highest performance over LTSA and other improved methods based on LTSA, and achieving the highest recognition rate in lower-dimensional embedding. The algorithm can effectively reconstruct low-dimensional coordination in curves with changed-density and also shows adaptability in high-dimensional images targeted at lower dimensions.

Keywords LTSA · ICA embedding · Principal tangent vectors · Local tangent space reconstruction

Q. Yang (✉) · F. Sun

State Key Laboratory of Intelligence Technology and Systems, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, People's Republic of China
e-mail: yangqw11@mails.tsinghua.edu.cn

Y. Li

School of Information Engineering, University of Science and Technology,
Beijing, People's Republic of China
e-mail: yuanli64@sina.com

Q. Yang

School of Electrical and Information Engineering, Beijing Jiao Tong University,
Beijing, People's Republic of China

1 Introduction

As the age of information technology proceeds, volume of information is increasing at an unprecedented speed typifying twenty-first century. Internet is loaded with images and videos that are growing at an exponential rate. Prophecy has seen the volume of information all over human civilization will be surpassed by that of the coming ten years. And for information, processing without dimensionality reduction will be unthinkable [1]. Many dimension reduction algorithms have been proposed, including linear subspace methods as PCA, LDA, classification techniques as SVM, clustering algorithm C-means, and kernel algorithm. Manifold learning was first proposed by J. B. Tenenbaum and S. T. Roweis with their corresponding publications in *Science* in 2000, in which Isomap [2] and LLE [3] were proposed. Some mainstream manifold learning algorithms are Isomap, LLE, Laplacian eigenmap, and LTSA [4].

One algorithm, LTSA, can detect the intrinsic dimension of manifolds, but is unreliable for unevenly distributed manifolds [4]. Scholars have proposed some algorithms to improve its performance. One is PLTSA [5] which can adapt to new samples well, but cannot solve the uneven sampling of manifolds. Another algorithm called LTSA+LDA [6] combines LTSA and LDA to conduct dimensionality reduction. Experiments on face databases show its higher recognition rate. However, in low-dimensional reduction, the algorithm cannot reconstruct information accurately.

In order to solve problems concerning higher-order information loss in LTSA, we propose the algorithm called ICA-embedded LTSA (IELTSA). The algorithm first searches local tangent space for tangent vectors. Then, ICA embedding is used to obtain independent components of the local tangent space for higher-order information. Finally, local space is aligned to reconstruct the manifolds, in which global low-dimensional mapping is obtained to represent the manifolds. Experiments are conducted on both synthetic curves and real-world face databases in both unsupervised and semi-supervised form. Results show that IELTSA is better than LTSA and other improved algorithms on unevenly distributed data and lower-dimensional embedding in face databases.

The rest of this paper is organized as follows: Section II introduces LTSA and ICA briefly. In section III, we discuss IELTSA algorithm by analyzing LTSA algorithm; furthermore, we prove the feasibility of the algorithm and analyze the reconstruction error. For section IV, IELTSA is used to reduce the dimensionality of 3D curves and also the UMIST face database. Finally, in section V, we analyze the algorithm, pointing out its advantages over other algorithms and concluding the paper with direction to enhance the performance of dimensionality reduction process.

2 Algorithms Introduction

2.1 Local Tangent Space Alignment

LTSA [4] is based on tangent space in manifolds and its central idea is to reconstruct geometric relationship with the help of local tangent space in the neighborhood of the samples. Then, the local tangent space alignment is used to calculate global embedding coordination of the manifolds. In this process, the transformation matrix can also be obtained for local embedding.

The central idea of local tangent space alignment method is to use the tangent space in the neighborhood of a data point to represent the local geometry and then align those local tangent spaces to construct the global coordinate system for the non-linear manifold. Compared with LLE and other manifold learning algorithms, LTSA is much faster and also adaptive to complex non-linear manifolds. But LTSA is sensitive to sampling density and data distribution, which makes it disadvantaged.

2.2 Independent Component Analysis

The central idea of independent component analysis is to find the intrinsic independent factor of non-Gauss distribution, especially for data which distribute as near-normal. ICA algorithm [7] is composed of preprocessing and model calculation. For preprocessing, centralization of data and whitening of data are involved. The result is centralized matrix

$$M = X - \bar{x} = X - \left(\sum_{i=1}^N y_i \right) / N \quad (1)$$

and this process will produce a whiten matrix V_d which is the first eigenvectors of covariance matrix $D(X, X) = (MM^T)/N$

And the result is $Y = V_d^T M$.

Then, the decomposition mapping matrix is calculated using mathematic model. In article [8], fast-ICA is introduced which can perform independent analysis by performing inversion via iteration. The recursive algorithm for calculating unmixing matrix is

$$\begin{cases} W_d^*(k) = D^{-1} E \left\{ Y (W_d(k-1)^T Y)^3 \right\} - 3W_d(k-1) \\ W_d(k) = W_d^*(k) / \sqrt{W_d^*(k)^T \times C \times W_d^*(k)} \end{cases} \quad (2)$$

where $C = (YY^T)/N$ is the covariance matrix of Y .

3 IELSTA Algorithm and Analysis

To facilitate the discussion of our proposed algorithm, we firstly explain the concept:

Definition Independent principal tangent vectors: suppose $L = \Theta \Sigma V^T$ and $U = \Theta_d \Sigma_d V_d^T$, where vectors of U are the first to d th largest left singular vectors; therefore, U is chosen as base vectors. It is called principal tangent space. Furthermore, if these chosen principal tangent spaces are independent as well, they are defined as independent tangent vectors.

3.1 Analysis of LTSA Algorithm

In LTSA algorithm [9], at the very beginning, we will illustrate that singular vector decomposition is used to ensure that the principal tangent vectors are chosen as base vectors in neighborhood $X_i = \{x_{i_1}, \dots, x_{i_k}\} \subset X \subset M$

Proof Suppose $X = ce^T + UT + E \subset M$

Firstly, neighborhood matrix is used to gain the local tangent space U_i using singular value decomposition,

$$X - \bar{x}e^T = \Theta \Sigma V^T \quad (3)$$

Set

$$\Sigma_d = \text{diag}(\sigma_1, \dots, \sigma_d), UT = \Theta_d \Sigma_d V_d^T, \Theta_d$$

and V_d are left and right singular matrixes for UT . Index σ_{d+1}/σ_1 is used as an auxiliary indicator in determining the reduced dimension of manifolds in LTSA, which satisfies the minimum reconstruction error:

$$\begin{aligned} \frac{\|E\|}{\|X\|} &\leq \frac{(\|X\| + \|ce^T + UT\|)}{\|X\|} = 1 - \frac{\|X - \Theta_d^c \Sigma_d^c (V_d^c)^T\|}{\|X\|} \\ &\approx \frac{|\theta_{d+1} \sigma_{d+1} V_{d+1}^T|}{\|\Theta \Sigma V^T + \bar{x}e^T\|} \quad (\text{1st order approximation}). \quad (4) \end{aligned}$$

Given $\sigma_{d+1}/\sigma_1 < \delta$, local reconstruction error:

$$\min \|E\| = \min_{c, U, T} \|X - (ce^T + UT)\|_F \quad (5)$$

which can be within a limit, by calculating the principal tangent vectors using SVD or generalized PCA.

However, alignment of LTSA is based actually on a local second-order and lower-order local information of tangent vectors, thus, making the global coordination unreliable with information losses, especially for the unevenly distributed space of its third-order convexity information. Thus, ICA-embedded LTSA is proposed as a non-linear embedding algorithm. In local neighborhood, we use ICA extraction, which can perform above second-order analysis, and thus, the global coordination can be reconstructed with higher-order information.

3.2 IELTSA and its Application

In this paper, IELTSA (ICA-embedded LTSA) is proposed to optimize the global coordination extraction using ICA embedding.

The algorithm is based on fully overlapped local tangent space which satisfies $f : O \subset \mathcal{R}^d \rightarrow \mathcal{M} \subset \mathcal{R}^n$ where function f is an isomorphic mapping.

Thus, by using independent component analysis to extract local information, we have to prove local reconstruction is possible and feasible.

Proof Singular vector decomposition method can get orthogonal local tangent space $\Theta_i = [\theta_1^{(i)}, \dots, \theta_{k_i}^{(i)}]$, satisfying

$$\theta_j^{(i)} = \theta_i^T (x_{ij} - \bar{x}_i) \quad (6)$$

where \bar{x}_i is the mean in X_i neighborhood.

In the independent component analysis,

$$\Lambda_i = \Upsilon_i \Theta_i, \Phi = \Upsilon_i^{-1} \quad (7)$$

where Υ_i is the unmixing matrix and Φ_i is the multiplex matrix.

Thus, ICA algorithm must be ensured to have no more reconstruction errors than embedding in LTSA.

Using IELTSA for tangent space alignment, local coordinates can be aligned [10] as

$$\tau_{ij} = \bar{\tau}_i + L_i \theta_j^{(i)} + \varepsilon_j^{(i)} \quad (8)$$

where L_i is the transform matrix, $\varepsilon_j^{(i)}$ local error. If $\theta_j^{(i)}$ is replaced by $\lambda_j^{(i)}$, note that $\lambda_j^{(i)} = \sum_{k=1}^n v_k \theta_k^{(i)}$, v_k is the projection of $\lambda_j^{(i)}$ onto $\theta_k^{(i)}$. According to equation (7),

$$\tau'_{ij} = \bar{\tau}'_i + L'_i \lambda_j^{(i)} + \varepsilon_j^{(i)'}, T'_j = \bar{T}' + L'_i \Lambda_i + E'_i \quad (9)$$

where L'_i is unknown; to find the solution for minimum error, we have the reconstruction error matrix:

$$E'_i = T'_i \left(I - E^{(i)} \right) - L'_i \Lambda_i \tag{10}$$

where $\mathcal{E}^{(i)} = ee^T/n$, e is vector of all ones and $\mathcal{E}^{(i)}$ is matrix of all ones.

On condition that T_i is determined, the optimal solution L_i for the objective

$$\min \|E'_i\|_F^2 = \left\| T'_i \left(I - \mathcal{E}^{(i)} \right) - L'_i \Lambda_i \right\| \tag{11}$$

Solution by MSE is

$$L'_i = T_i \left(I - \mathcal{E}^{(i)} \right) \Lambda_i^{+\alpha} \tag{12}$$

Minimum error is

$$E'_i = T'_i \left(I - \mathcal{E}^{(i)} \right) \left(I - \Lambda_i^+ \Lambda_i \right) \tag{13}$$

Similarly, for LTSA, by the principal tangent vector, the optimal solution is

$$L_i = T_i \left(I - \mathcal{E}^{(i)} \right) \left(I - \Theta_i^+ \Theta_i \right) \tag{14}$$

Minimum error

$$E_i = T_i \left(I - \mathcal{E}^{(i)} \right) \left(I - \Theta_i^+ \Theta_i \right) \tag{15}$$

Easy to prove that the difference between their reconstruction error is of second order in terms of local coordination $\|E_i - E'_i\|_F \sim o\left((\tau_i - \bar{\tau})^2\right)$, that is,

$$\begin{aligned} & \left\| T_i \left(I - \mathcal{E}^{(i)} \right) \left(I - \Theta_i^+ \Theta_i \right) - T'_i \left(I - \mathcal{E}^{(i)} \right) \left(I - \Lambda_i^+ \Lambda_i \right) \right\|_F^2 \\ &= \left\| I - \mathcal{E}^{(i)} \right\|_F^2 \cdot \left\| T_i \left(I - \Theta_i^+ \Theta_i \right) - T'_i \left(I - \mathcal{E}^{(i)} \right) \left(I - \Lambda_i^+ \Lambda_i \right) \right\|_F^2 \\ &= C \left\| T_i \left(I - \left(\Phi_i \Lambda_i \right)^+ \Phi_i \Lambda_i \right) - T'_i \left(I - \Lambda_i^+ \Lambda_i \right) \right\|_F^2 = C' \left\| T_i - T'_i \right\|_F^2 \sim o\left(\left(\tau_i - \tau'_i \right)^4 \right) \\ & \left(\text{for } d\left(\tau_{i_p}, \tau_{i_q} \right) \doteq d_M\left(x_{i_p}, x_{i_q} \right) \doteq d\left(\tau'_{i_p}, \tau'_{i_q} \right) \right) \end{aligned}$$

As LTSA can approximate the manifolds in $o\left(\|\bar{\tau} - \tau\|^2\right)$, IELTSA has equal-order reconstruction error to approximate manifolds in lower dimensions.

Furthermore, the transformation matrix T'_i is the eigenvectors of $B = SW'W'TS^T$ corresponding to its second to $(d + 1)$ th smallest eigenvalue, where $W' = \text{diag}(W'_1, \dots, W'_n)$, $W'_i = \left(I - \mathcal{E}^{(i)} \right) \left(I - \Lambda_i^+ \Lambda_i \right)$ and $S = (S_1, \dots, S_n)$, S_i is the (0–1) selection matrix. Furthermore, by the mixing matrix Φ and Eqs. (12–14), we can obtain

$$L_i = L'_i \Phi_i^{-1} \quad (16)$$

Compared with LTSA, the transformation matrix $L_i = L'_i \Phi_i^{-1} = T_i (I - \mathcal{E}^{(i)}) \Lambda_i^+ \Phi_i^{-1}$ is combined with mixing matrix Φ ; therefore, this makes it dual-dependent on ICA process. For this reason, we introduce the factor α as a normalization factor in ICA-embedded algorithm to regulate the tangent vectors. After normalization, we will get V_i as the final tangent space.

Now, we have extracted local information and proved the availability of reconstruction method, and then, we can calculate the alignment matrix Ψ to reconstruct the global coordination. The results are transformation matrix L_i and lower-dimensional embedding $mappedX \subset R^a$ is obtained.

3.3 IELTSA Algorithm

Input: X , classification labels, target dimension d , normalization factor α .

Output: Lower coordination embedding $mappedX$, local unmixing matrix Υ , or local mixing matrix Φ .

Step1: Local Neighborhood Construction: For data set $X = \{x_1, \dots, x_n \in R^d\}$, each x_i has its k -nearest neighbors to construct neighborhood matrix $X_i = \{x_{i_1}, \dots, x_{i_k}\}$;

Step2: Construct local coordination: To approximate neighborhood information of x_i , firstly calculate local coordination $\theta_1^{(i)}, \dots, \theta_k^{(i)} \in R^d$ by SVD, which is d largest left singular eigenvectors of the neighborhood matrix X_i , satisfying

Step3: Local ICA embedding: For dimensionality d , independent component analysis is applied to local tangent space, making sure that the interconnected information is minimum, that is, independent tangent vectors will construct a tangent space $\Lambda_i = \Upsilon_i \Theta_i$, where Υ_i is local mixing matrix;

Step4: Normalized space calculation: with normalization factor α , the final local tangent space is V_i .

Step5: Transformation matrix: By calculating alignment matrix Ψ_i , lower-dimensional embedding is reconstructed and for neighborhood of each x_i , transformation matrix is calculated by $L'_i = T_i (I - \mathcal{E}^{(i)}) V_i^+$

Step6: Coordination reconstruction: By calculating

$$\min \|E_i\|_F^2 = \left\| T_i (I - \mathcal{E}^{(i)}) (I - V_i^+ V_i) \right\|_F^2$$

final embedding $T_i = [\tau_{i_1}, \dots, \tau_{i_D}] \in R^{d \times D}$ with orthogonal column vectors and indexes i_1, \dots, i_D are determined by neighborhood of x_i

4 Experiments and Analysis

Experiments are carried out on Intel Pentium D 2.78 GHz, 2.0 GB memory and mainly on Matlab R2010a software. Experiments are performed on the proposed IETLSA and other algorithms including LTSA and improved methods PLTSA and LTSA+LDA to make a comparison.

4.1 Unevenly Distributed Curves

For the local degraded Swiss-roll, experiments employ four different algorithms to show their adaptability.

Figure 1b illustrates the results of the four algorithms on the local degraded Swiss-roll. PLTSA and LTSA+LDA results are highly overlapped. For LTSA, the unevenly distributed area is distorted and makes a “rolled tail,” thus cannot reflect the manifold accurately. However, IELTSA adapts to the “tail” quite well and extract the local area in a sorted manner, thus performs better than other methods within error. The results corroborate the higher-order analytical ability of ICA. Experiments prove that it can adapt to uneven distribution of data.

4.2 UMIST Face Database

UMIST face database (website: <http://images.ee.umist.ac.uk/danny/database.html>) consists of 564 pictures from 20 persons of different postures, genders, races, and ages as well. There are around 19–36 pictures for each person. It has pictures of different angles from 90° profile to frontal images. Figure 2 shows some samples of this database. We apply algorithm IELTSA and other improved algorithms to UMIST database. For unsupervised dimensionality reduction, we use LTSA, PLTSA, and LTSA+LDA for comparisons with our proposed algorithm. In

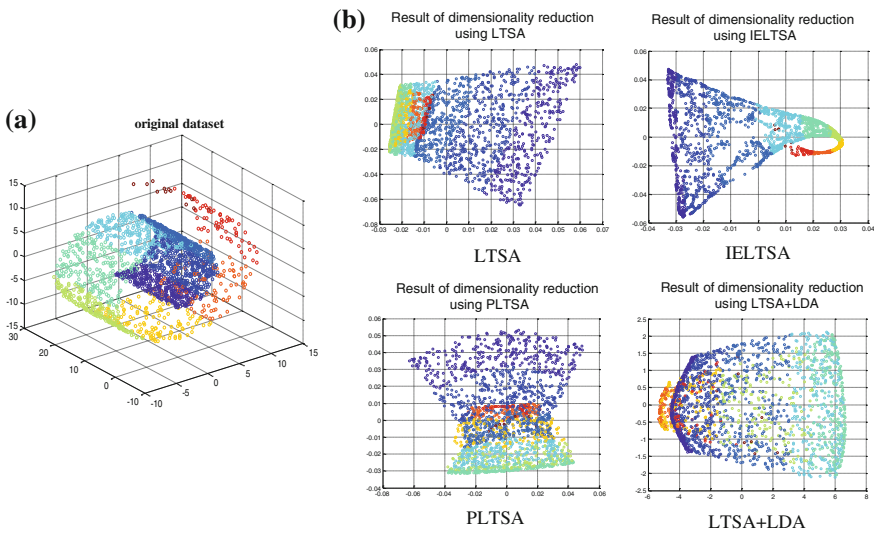


Fig. 1 Result on uneven distributed Swiss-roll **a** original data set; **b** the results of different algorithms



Fig. 2 Samples of UMIST face database

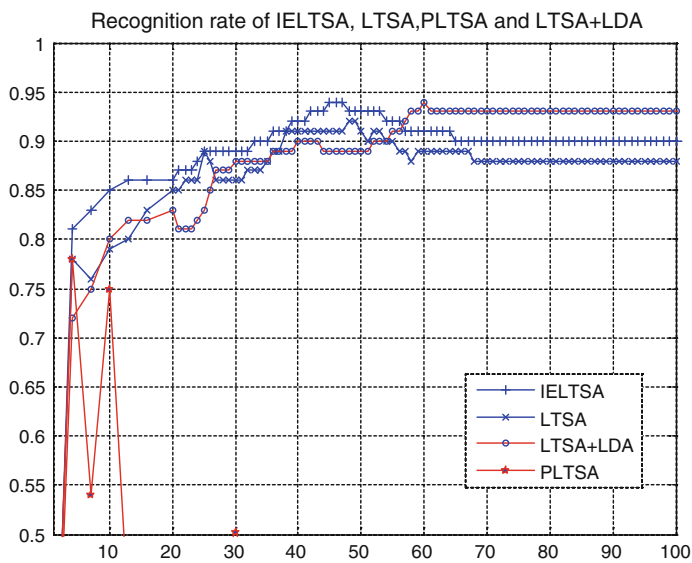


Fig. 3 Recognition rate as a function of dimensionality

order to get visual results, we obtain 2D mapping results by projecting onto $x - y$ plane as shown in Fig. 3.

In analyzing Fig. 3, to compare LTSA with the proposed algorithm, IELTSA shows classification in lower-dimensional mapping better, due to the independent tangent vectors extracted via local information calculation, which can facilitate the classification.

Then, we conduct experiments on a semi-supervised learning with randomly chosen 100 samples (5 images per person) as test set and the remaining for training. The predefined labels are used as supervision on which k -nearest neighbors is calculated (Set $k = 11$). Experiments with target dimension varying from 4 to 100 are conducted for 15 iterations each. Figure 4 shows the recognition rate as a function of targeted dimension. Recognition rate of each algorithm with different dimensionality is rounded to the nearest ones. Then, for the best recognition rate of each algorithm, different test sets are experimented on for 50 iterations and Table 1 shows the average recognition rate.

Figure 4 and Table 1 show that IELTSA can reduce the dimensionality varying from relatively low to high dimensions. For high-dimensional data which must be reduced to restricted dimension, IELTSA shows greater advantage over other

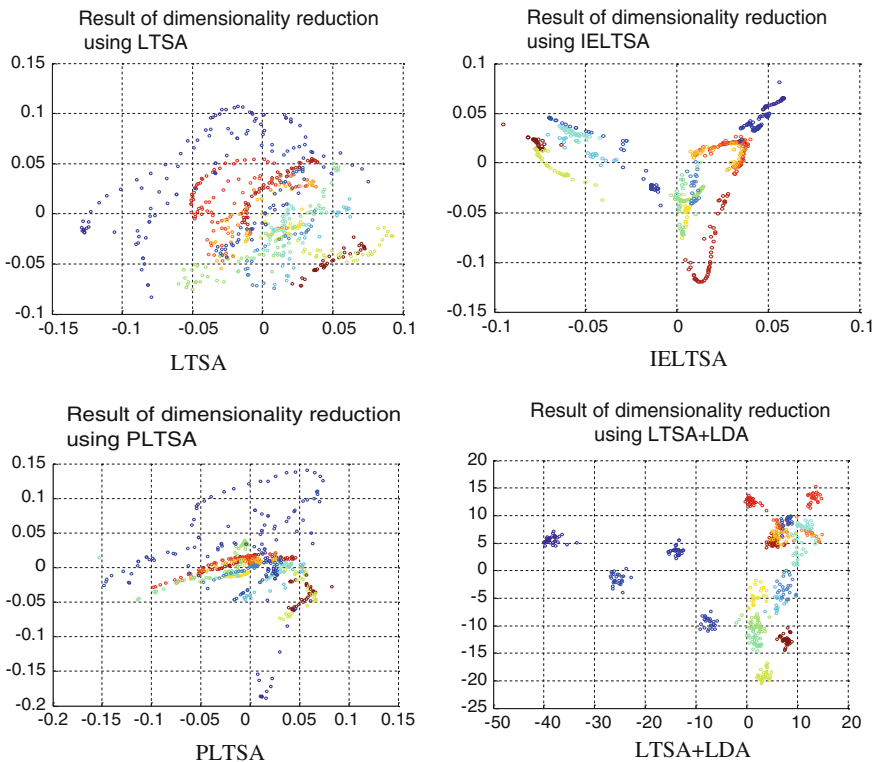


Fig. 4 Results of dimensionality reduction in UMIST

Table 1 Recognition rate for UMIST face database

Method	LTSA	IELTSA	LTSA+LDA
D_{opt}	50	45	≈ 60
R_{avg}	92.66	94.40	94.78

D_{opt} Optimal recognition dimension

R_{avg} Average recognition rate

algorithms. LTSA can adapt to face recognition in UMIST. But the recognition rate is lower compared with other improved methods based on LTSA. PLTSA is not as adaptive as other techniques perhaps because linear analysis performed on manifolds will lead to loss of critical information in high dimension. LTSA+LDA is very effective when target dimensionality is relatively high; however, in lower dimension, it has lower performance than IELTSA. Experiments show that compared with LTSA and other improved algorithms, IELTSA advantages in stability with regards to target dimension. Also, IELTSA can apply to both synthetic data and natural-state face databases and show advantages over other algorithms.

In sum, ICA-embedded LTSA is an effective algorithm, which can not only perform dimension reduction better in 3D curves, but also perform face image reduction in lower dimensionality better than other methods.

5 Conclusion

Experiments and analysis show that algorithm IELTSA is not only feasible theoretically, but also adaptive to synthetic and natural datasets as manifold learning algorithm. IELTSA shows better performance for unevenly distributed data than LTSA and other improved algorithms. Also, it shows robustness on target dimension in face database and also achieves good results in lower-dimensional mapping.

However, the proposed algorithm still has deficiency, and some aspects need to be further improved.

Firstly, complexity analysis is needed upon which higher dimensionality reduction is more dependent. With the introduction of ICA embedding, both spatial and temporal complexity should be discussed. Secondly, application of IELTSA in image processing should be explored further. Future work will be to test the use of IELTSA in pattern recognition and other fields, and research into higher dimension to raise recognition rate. Finally, the normalization factor we introduce has normally set to one but it is critical in higher dimension reduction, making the proper parameter setting necessary. Theoretical analysis and experiments are both needed in which we will discuss this factor further.

References

1. Zhu T (2009) Manifold learning and application in image processing, Dissertation, Beijing Jiaotong University
2. Tenebaum JB, Silvam VD, Longford JC (2000) A global geometric framework for nonlinear dimensionality reduction. *Science* 290:2319–2323
3. Rowels ST, Saul LK (2000) Nonlinear dimensionality reduction by locally linear embedding. *Science* 290:2323–2326
4. Zhang Z, Zha H (2004) Principal manifolds and nonlinear dimension reduction via local tangent space alignment. *SIAM J Sci Comput* 26:313–338
5. Yang J, Li FX, Wang J (2005) A better scaled local tangent space alignment algorithm. *J Softw* 16:1584–1590. doi:[10.1360/jos161584](https://doi.org/10.1360/jos161584)
6. Zhou C, WEI X, Zhang Q, Bai C (2009) Face recognition based on ICA and features fusion. *J Basic Sci Eng* 17:799–812. doi:[10.3969/j.j.issn.1005.0930.2009.05.0019](https://doi.org/10.3969/j.j.issn.1005.0930.2009.05.0019)
7. Dun W, Mu Z (2009) An ICA-based ear recognition method through nonlinear adaptive feature fusion. *Comput Aided Design Comput Graph* 21:382–390
8. Hyvarinen A (1999) Fast and robust fixed-point algorithm for independent component analysis. *IEEE Trans Neural Netw* 10(3):624–634
9. Ye Q, Zha H, Li R (2007) Analysis of an alignment algorithm for nonlinear dimensionality reduction. *BIT Numer Math* 47:873–885. doi:[10.1007/s10543-007-0144-x](https://doi.org/10.1007/s10543-007-0144-x)
10. Chen WH (1998) Differential geometry preliminary. Higher Education Press, Beijing

Analysis and Improvement of Anonymous Authentication Protocol for Low-Cost RFID Systems

Zhijun Ge and Yongsheng Hao

Abstract With the rapid growth of RFID applications, security has become an important issue. Security protocols work as a kernel for RFID technology. In this paper, we propose a serverless protocol to protect the system from suffering different type of attacks. Moreover, safety requirements for RFID protocols were analyzed, and a low-cost anonymous authentication protocol for RFID was proposed based on the universal composability mode. This protocol is feasible under a desynchronizing attack by recovering the disabled tags that are desynchronized with the reader because of this attack. The improvement of our scheme is the index item used for an advanced search. Finally, sufficient analysis is given to prove security quality of the protocol.

Keywords Security · Authentication protocol · Desynchronizing attack · Radio frequency identification (RFID)

1 Introduction

Radio frequency identification (RFID), as an automatic identification technology, is developed with the maturity of communication technology and large-scale integrated circuits. RFID tags are readable without line-of-sight contact, with characteristics of large data storage capacity and environmental adaptability. Accounting to its own advantages, RFID technology can be widely used in the fields of traffic control, supply chain, disease traceability, and waste collection, etc. [1].

Z. Ge (✉) · Y. Hao
Department of Navigation Engineering, Mechanism Engineering College,
Shijiazhuang, China
e-mail: abcdefgewhut@gmail.com

In the typical architecture of RFID systems, a tag has independent memory and computing form, which can achieve the access control and encryption. The reader broadcasts an RF signal and forwards the tag's response to the backend server. The server can retrieve the detailed information of the tag only after identifying the tag successfully.

RFID technology may improve efficiency and reducing costs for the business or organization [2]. However, the security and privacy problems of RFID systems, such as eavesdropping, cloning, impersonations, and tracking of the users, may confine the popularization of RFID technology, and RFID security issue has attracted more and more attention to propose efficient ways and means. There are two important factors that impact the function of security technique [2–4]: Firstly, RFID tag has very limited power, storage capacity, and computing ability. Secondly, the communication signals between the tag and reader are transmitted by wireless channel, which is exposed to attackers.

Authentication can be one approach to address such security and privacy threats. Several authentication protocols have been proposed in past years. As the backend server can provide an easy way for checking the validity between the tag and the reader, most of the works draw their considerations on a server-based model. There are many significant symmetric key protocols proposed for RFID security and privacy, just as hash-lock [3], the randomized hash-lock [4], OSK protocol [5]. Tan et al. [6] propose an approach to solve the problem when a safe and continuous connection cannot be charged between the reader and the backend server. In such serverless system, a reader has to identify legitimate tags all by itself because of the absence of server. Meanwhile, he also has to make a self-verification to receive the tag's data.

In fact, further issues confuse the serverless system badly. Portable readers can be stolen like tags, and authentication protocols without anonymity may lead to be a severe breach in the security of an RFID system. On the other hand, the limited resources of mobile readers are in a big trouble to recover from unexpected conditions during operation. This paper proposes an authentication scheme that can provide security and privacy protection similar to the backend server model without having a persistent connection. To ensure robustness of the system, our approach aims to address anonymity and can get back from desynchronizing attack [1, 6–8]. Including the communication status would have a better protection.

The rest of paper is organized as follows: Sect. 2 gives the preliminary works for an anonymous authentication protocol. The proposed protocol will be described in Sect. 3. Section 4 defines the attack model and analyzes the security of the protocol. We present the performance analysis of the protocol in Sect. 5. Finally, we conclude our works in Sect. 6.

2 Preliminaries

Our serverless RFID system has two major components: the reader and a set of tags. In order to get enough security indicators, Tan et al. [6] use the backend database as a certification authority (CA). CA initializes each tag and reader by writing the secret into memory, so that authorized their access to each other.

In this paper, we get inspirations from Tan et al. scheme, we build a similar certification authority (CA), for each legitimate reader has to download certified lists from CA as a certificate. The channel between readers and the CA is assumed to be robust enough, whereas attacks are assumed to happen in the communication channel between the reader and the tag.

Assume that each tag and each reader have the knowledge of some specified irreversible one-way functions, just as hash function, the pseudorandom number generator (PRNG). by which a current seed cannot be linked with the previous one. In [9], Ma provides a comparatively rigorous proof of that tag generates pseudorandom number is the necessary condition to guarantee RFID forward privacy. We choose a fairly simple function $P(.)$ to generate pseudorandom numbers as he can be implemented at lower cost.

$P(.)$ takes an argument and outputs a pseudorandom number according to its distribution. Suppose that there are n tags in the system, and each tag is initialized with a secret number S and a unique identifier ID by CA. On the other hand, the reader obtains his identifier r and a contact list L from the CA during the system deployment. The contact list L contains information about the tags that composed of the secret number S and the secret number of the last successful session S_{pre} , their corresponded keys I and I_{pre} by all appearances. These keys allow the unequivocal identification of the tag and can be used as an index to allocate all the additional information linked to the tag. In other words,

$$L = (I, I_{pre}) \begin{cases} S_1, S_{pre,1} \\ S_2, S_{pre,2} \\ \dots \\ S_n, S_{pre,n} \end{cases} \tag{1}$$

Now, we discuss the initial notation. Select a pseudorandom function $P: \{0, 1\}^l \rightarrow \{0, 1\}^{2l}$, $P(S) = P_1(S) || P_2(S)$ where $||$ represents concatenation and $|P_1(S)| = |P_2(S)| = l$. Secret value is shared between the tag and the reader, initially to the reader, $S_{pre,i} = S_i = S$, $I = I_{pre} = P_1(S)$, $\forall i \in [1, n]$. Note that the reader does not know the tag unique identifier ID, and the reader and tag can never learn ID or r each other from the received message.

3 Protocol Description

The protocol operates as shown in Fig. 1. The concrete process is described as follows:

At the beginning, the reader sends request and a random number R_r ; then, the tag responses with the message $\langle I, M, R_t \rangle$ for authenticating itself. R_t is another random number generated by the tag, and I and M are computed as follows:

$$S' = I || I' = P(S) = P_1(S) || P_2(S) \tag{2}$$

$$I \leftarrow P_1(S) \tag{3}$$

$$M \leftarrow P_1(S \oplus (R_r || R_t)) \tag{4}$$

The reader uses the index information I to check the validity of a tag. In most cases, the reader finds a match and becomes sure about the validity of the tag. For $I_i = I$, it just means the tag secret number has been updated then will find the tuple structured data $\langle S_i, S_{pre,i}, ID_i \rangle$. The reader calculates $P_1(S_i \oplus (R_r || R_t))$ to compare with M , whenever the tag is authenticated, it updates $S_{pre,i}$ with the current S_i , then mutates S_i to $P(S_i)$. The index keys update in a same way. For $I_{pre,j} = I$, it just means the tag has no updated pseudonym, it means that we should check and find the index key I_{pre} to match with tuple $\langle S_j, S_{pre,j}, ID_j \rangle$, the reader computes $P_1(S_{pre} \oplus (R_r || R_t))$ and make a comparison. After identifying the tag, the reader

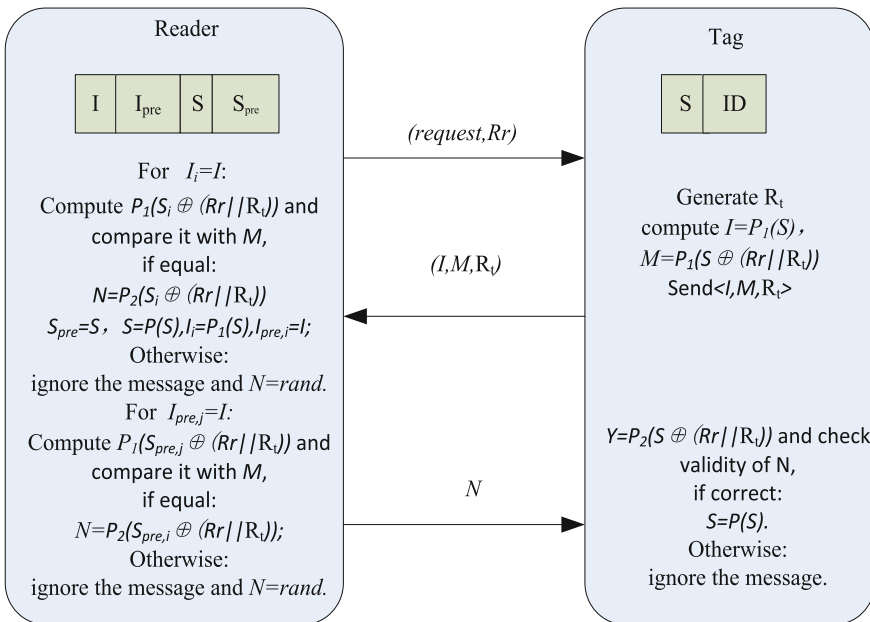


Fig. 1 The proposed protocol

has to prove itself to the queried tag. The reader has to reply N to the tag, which N is used for authentication.

$$N \leftarrow P_2(S_{\text{pre},i} \oplus (R_r || R_t)) \quad (5)$$

In all cases, the reader replies with the produced N . If the reader fails to find a match, it ignores the message and replies with a random number $N = \text{rand}$, and we conclude the tag as a fake tag. Next step is the tag's turn to verify validity of the reader. The tag produces pseudorandom number $Y = P_2(S \oplus (R_r || R_t))$. If the tag finds a match between Y and N , the tag updates his secret number accordingly with a pseudorandom number $P(S)$ and concludes the reader as an authorized one. Otherwise, the tag will discard the message and kill the session.

Both the reader and the tag update their secret number after they confirm the validity of the opposing side.

4 Parameter Setting

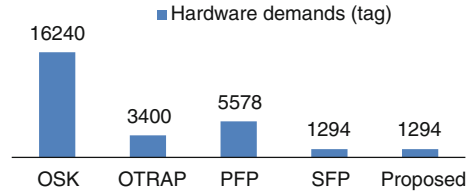
Our authentication protocol mainly involves one pseudorandom function $P(\cdot)$ and other combinatorial logic unit. But the cost depends on the number of execution of irreversible one-way functions during an authenticated session. According to [10, 11], the moderate weight security of RFID system is needed to achieve the security level of 80 bit. We can attain the expected purpose by using an algorithm of stream ciphers such as Grain and Trivium to construct a PRNG. Hardware demands will cost 3360 or 3090 gate equivalents (GE), respectively. In fact, the quantity of GE contains the occupied block by the initial vectors among those costs. However, RFID tags have no use of these initial vectors. Thus, we can construct a PRNG in our protocol with a simple structure, and the required number of gates can be reduced to 1294 GE [12].

The scheme [11] chooses the universal hash function based on the Toeplitz matrix. According to [13], the universal hash function requires about 4284 GE, which provides an 80-bit level of safety. For lightweight protocol, the required number of gates can be reduced to 1700 GE. In [14] introduces the SHA-1-based hash function, which can be constructed with approximately 8120 GE. Symmetric key schemes with the algorithm of AES-128 mentioned in [15] cost about 3400 GE.

We compare our protocol with some RFID authentication protocols suggested in recent years [5, 9, 12, 15]. This comparison is based on the hardware demands of a tag, which is shown as Fig. 2.

Our protocol is suggested as a serverless authentication protocol which ensures mutual authentication of both the tag and the reader. We charge the RFID reader with almost all the computation during the transaction of protocol.

Fig. 2 Comparison of hardware demands (Tag)



5 Security Analysis

We analyze our proposed protocol about the security quality here. We define an attack model firstly. Then, we focus on security analysis of the protocol.

5.1 Attack Model

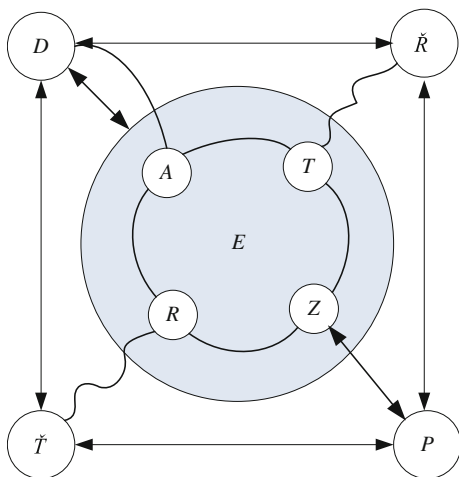
The major goal of an adversary in any RFID system can be described in commonly, fake tag is supposed to be counterfeit a real tag, such that he can be identified as a legitimate one, or adversarial reader puzzles the real tags to regard him as an authorized one.

In this section, an adversary is denoted as A based on the universal composability mode. Deng et al. [16, 17] construct a hit of emulator E in an ideal process. Our assumption includes that in any environment simulator Z , the adversary A can join in the protocol and cannot classify from the legitimate entities. The emulator S includes all adversarial entries. The protocol of the system is denoted as a black box P . The environment simulator Z can recognize all the imports and acquire their outputs at the same time. Moreover, Z directs all the entries about corresponding executions.

The emulator E simulates the adversary A communicates to legitimate reader R and tag T in the environment Z . In addition, E is also preserved a database D to record a temporary key for each tag. Correspond to the valid reader and tag, the adversarial reader and tag are denoted as \check{R} and \check{T} . \check{R} is unauthorized to have access to any real tags as it is not connected with the backend server. Similarly, \check{T} is not valid as it has no idea about real S and real ID. We presume that the backend server cannot be compromised; otherwise, the adversary would get total control over the tag database. We do not take for granted that the adversary has unlimited resources. Instead, we assume that all the entities such as tags, readers, adversaries, adversarial tags, and adversarial readers have polynomial bounded resources. Definitions and constructions of emulator E are shown as Fig. 3.

Theoretically, the adversary A can be supposed as a probability polynomial time (PPT) algorithm with powerful but bounded resources [9, 17]. He can completely control the communications between the legitimate reader R and the tag T in the environment Z . The adversary A in behaviors is described into oracle machine as follow:

Fig. 3 The construction of emulatur E



Launch(R). The adversary A launches the adversarial reader \check{R} to generate a new session of the protocol processing. With the output of the session identifier $ssid$ and the first round message m_1 , he simulates to eavesdrop the first session of the protocol processing.

SendTag($ssid, m_1, T_i$). It means that simulations of tampering message of the protocol, or the tag calculation $TagComputer(ssid, m_1, T_i)$. After all, send m_2 as the second round message.

SendReader($ssid, m_2$). Simulate to eavesdrop and tamper the message of the protocol processing, including the reader calculation $ReaderComputer(ssid, m_2)$. Then, he generates an output with the third rand message m_3 .

Reveal(T_i). Crack attacks to the tag T are simulated. In addition, A creates output with the internal state of the tag T .

Suppose that times of query oracle machine by the adversary A are limited to be q .

Accounting to the alternation of Z and the legitimate R and T , the emulatur E simulated as following in steps.

Once receiving a message from the protocol P , the emulatur E acquires the type of Cmessage (reader or tag) by function $Type()$. Then, write a table in structure of $Establish(ssid, Reader, list)$ or $Establish(ssid', Tag, list)$. For $Establish(ssid', Tag, list)$, if $list$ is $NULL$, it generates a key (S, ID) , which

$$S \leftarrow \{0, 1\}^k \tag{6}$$

$$ID \leftarrow \{0, 1\}^k \tag{7}$$

Otherwise, copy the key from the $list$.

Once the fake tag receives the message m , the emulatur E will order the environment Z to send m to T . If the valid tag T answers a message n , E shepherd \check{T} clone the message n to Z .

Despite of these actions, the adversary can block any channel at any time to full fill his purpose. The adversary can launch the desynchronizing attack by blocking (or jamming) any of the channels at any step of the protocol or by scrambling any message passed from one party to another.

5.2 Security Analysis

In this subsection, we analyze our proposed protocol under the attack model defined above and explain how our protocol protects the system from suffering different type of attacks.

In our attack model, all the querying information is transmitted by the environment Z controlled by the emulatur E . The adversary A can eavesdrop and record all the information. Furthermore, the database D provides a data supporting to establish a link between the responses. The adversarial reader \check{R} and tag \check{T} are trying to join the protocol as possible.

Privacy protection. The static identifier of the tag is never sent to anyone, not even to the authorized reader. All the communications between the reader and the tag are produced in Boolean calculation, specifically. When adversary A queries the legitimate tag T , it replies to the reader with new response $M \leftarrow P_1(S \oplus (R_r || R_t))$. Because of the irreversible nature of PRNG, the adversary A cannot get valid information from the responses. Thus, our protocol protects the privacy of the tag.

Resist replay attacks. The adversary A can eavesdrop all messages transmit between the reader and the tag, just as M and N et al. By incorporating random numbers R_r , R_t or $rand$, our protocol becomes secured against replay attacks, since each session has the same random number probability $1/2^l$, in which l is the length of the random numbers. Accordingly, the probability of successful certification can be ignored when an attacker just replay the preterit information.

Prevent Tracking. The adversary A tries to track T over time. While tracking, the adversarial reader \check{R} can reuse a same random number $rand'$ learned from any previous session of our protocol. By incorporating $rand'$ in every session, our protocol becomes robust enough to fight against tracking since the emulatur E cannot predict $rand$ ahead of time. Even the number $N = rand$ keeps the protocol consistent by preventing emulatur E to acquire any knowledge (success or failure) about this session, that is, she cannot track T afterward.

Ensure Forward Secrecy. Forward secrecy means that an adversary cannot realize any previous information by an entity even if he compromises it [15, 18]. If the emulatur E get any key heuristic information of tags, just as the unique identifier ID, the adversary A may trace the data back through past every information exchanged between R and T in which messages had been recorded at database D . Our protocol ensures forward secrecy since responses based on the irreversible one-way function $P(\cdot)$. The former secret numbers are used as an argument of function $P(\cdot)$, and a current seed cannot be linked with the previous one.

Lessen Susceptibility to Desynchronizing attack. During a successful tag query, the reader and the tag are both in sync. In fact, the emulator E can break the synchronicity [19] in final communication pass of the protocol by jamming the environment Z , while T waits for the message N from R . Since the last message N is damaged or lost, the tag cannot update his secret number, he becomes desynchronized with the real reader. Our protocol provides the previous information to counterwork against the desynchronizing attack. When we fail to find a match by search the current data, the reader R will further continue the search with the previous secret numbers, I_{pre} and $S_{pre,i}$. For legitimate tag T , the search will be fruitful since the secret number S never updated before T confirm the validity of the opposing side. At this step, readers can recover the valid tag from desynchronizing state.

6 Conclusions

In this paper, safety requirements for RFID protocols were analyzed, and a low-cost anonymous authentication protocol for RFID was proposed based on the universal composability mode. The implementation of this protocol is feasible for a wide range of RFID architectures. Our protocol provides a protection against some major attacks, just as tracking, replaying and desynchronize attack, which can also ensure the forward secrecy of system.

Under our protocol, the provisional random number generated by the Protocol Label secret transmission index value, the active attacker can prevent the tracking label, so reducing the efficiency of a fixed index value in the backend database search; on the other hand, the proposed protocol avoid the logic exhaustive search in the backend database. Contrarily, we replaced it by a repeated-bit computing since special equipment with considerable resources is necessary to mount them.

References

1. Juels A (2006) RFID security and privacy: a research survey. *IEEE J Sel Areas in Commun* vol 24(2), pp 381–395
2. Kerschbaum F, Sorniotti A (2009) RFID-based supply chain partner authentication and key agreement. In: *ACM conference on wireless network security (WiSec 09)*. pp 41–50
3. Weis S (2003) Security and privacy in radio frequency identification device. MIT, Cambridge
4. Weis S, Sarma S, Rivest R, Engels D (2003) Security and privacy aspects of low-cost radio frequency identification systems. In: *international conference on security in pervasive computing (SPC03)*, pp 454–469
5. Ohkubo M, Suzuki K, Kinoshita S (2003) Cryptographic approach to “Privacy-Friendly” tags. In: *RFID privacy workshop*, pp 624–654. USA
6. Tan C, Sheng B, Li Q (2007) Serverless search and authentication protocols for RFID. In: *annual IEEE international conference on pervasive computing and communications (PerCom 07)*, pp 3–12. IEEE Press, New York

7. Ahamed I, Rahman F, Hoque M, et al (2008) YA-SRAP: Yet another serverless RFID authentication protocol. In: IET international conference on intelligent environment (IE 08), pp 1–8. IEEE Press, New York
8. Hoque M, Rahman F, Ahamed S, Park J (2009) Enhancing privacy and security of RFID system with serverless authentication and search protocols in pervasive environments. Springer wireless personal communication
9. Ma CS (2011) Low cost RFID authentication protocol with forward privacy. Chin J Comput. China, vol 34 Aug, pp 1387–1398
10. Luo L, Chan T, Li JS et al. (2006) Experimental analysis of an RFID security protocol. In: IEEE international conference on e-Business engineering (ICEE 06), pp 62–70
11. Feldhofer M (2007) Comparison of low-power implementations of Trivium and Grain. In: workshop on the state of the art of stream ciphers (SASC 07), pp 236–246
12. Haitner I, Reingold O, Vadhan S (2010) Efficiency improvements in constructing pseudorandom generator from any one-way function. In: ACM symposium on theory of computing (STOC 10), pp 437–446
13. Yksel J, Kaps JP, Sunar B (2004) Universal hash functions for emerging ultra-low-power networks. In CNDS
14. Feldhofer M, Wolkerstoefer J (2007) Strong crypto for RFID tags—A comparison of low-power hardware implementations. In: IEEE international symposium on circuits and systems (ISCAS 07), pp 27–30
15. Berbain C, Billet O, Etrog J et al. (2009) An efficient forward private RFID protocol. In: 16th ACM conference on computer and communications security (ACM CCS' 09), pp 43–53. Chicago, USA
16. Miaolei D, Jianfeng M, Fulong L (2009) Universally composable three party password-based key exchange protocol. China communications, vol 6(3), pp 150–155
17. Deng ML, Wang YL, Qiu G et al. (2009) Authentication Protocol for RFID without back-end database. J Beijing Univ Posts Telecommun, vol 32(4), pp 59–62
18. Conti M, Pietro R, Mancini L, et al. (2007) RIPP-FS: an RFID identification, privacy preserving protocol with forward secrecy. In: IEEE international workshop on pervasive computing and communication security (PCCS 07), pp 229–234 IEEE Press
19. Hoque M, Rahman F, Ahamed S (2009) Supporting recovery, privacy and security in RFID systems using a robust authentication protocol. In: 24th ACM symposium on applied computing (ACMSAC 09), pp 1062–1066

Formation Control and Stability Analysis of Spacecraft: An Energy Concept–Based Approach

Zhijie Gao, Fuchun Sun, Tieding Guo, Haibo Min and Dongfang Yang

Abstract In this paper, the spacecraft formation control and its stability problems are studied on the basis of the energy concept. The formation systems are viewed as a multi-mass point system with both generalized elastic deformation and rigid body movement, and certain potential fields interact with each other within the formation. Consequently, the stability of formation and coordination and the controller design problems are studied from the perspectives of energy. The symmetry in formation motion is studied, and the definition of formation stability and its corresponding criteria are presented based upon the notion of relative equilibrium. Then, the artificial potential method is explored to design the formation control law, and the stability of which is followed by utilizing the former criterions. The effectiveness of the proposed formation control approach is also demonstrated by numerical results.

Keywords Formation control · Energy concept · Stability analysis · Spacecraft

1 Introduction

The past few years have witnessed the burgeoning interest in spacecraft formation control. In spacecraft formation systems, each spacecraft works cooperatively to complete more complicated tasks which are unavailable by single spacecraft, and

Z. Gao (✉) · F. Sun · T. Guo · H. Min
Tsinghua University, Beijing, China
e-mail: gzjie@163.com

D. Yang
Hi-Tech institute of Xi'an, Xi'an, China
e-mail: ydf09@mails.tsinghua.edu.cn

this cooperation can improve the functionality, flexibility and the reliability of the system. In spacecraft formation systems, formation control is an essential and enabling technique. In most of the previous studies, the formation control problem is directly transformed into a motion tracking problem based upon linearized motion formulation [1]. And the corresponding controller can be designed through linear system theory. It is worth noting that the linearized motion formulation-based control method is difficult to give a natural description from the physical concept; thus, it is difficult to formulate the formation movement as a whole. In addition, although most literatures on formation flying control involve the formation stability problem [2–6], few of them study the overall movement from the perspectives of the stability of whole team.

Energy-based control is one of the two basic frameworks of control theory [7, 9–11]. For dynamic systems, the concept of energy is more reasonable. The extreme points of the potential energy always correspond to the balance point, and the stability can be easily determined by calculating the positive definiteness of the potential energy function near the equilibrium point. In this paper, the theory of dynamical systems and its energy concept is utilized to study the formation stability and its control problem. The formation movement is decomposed into the deformation and the overall movement of the formation, which reflects the symmetry of the formation movement. First, the Jacobi coordinates and shape coordinates are introduced to eliminate the rigid motion degree of freedom (DoF). Based upon this deforming movement formulation and relative balance concept in geometric mechanics [3], this paper presents strict mathematical definitions of balance formation and stable formation. In addition, we also use the artificial potential technology to design the energy mode that corresponds to the desired formation kinetics [4, 6, 7]. From this energy mode, we can obtain the desired formation control law by back stepping methods. Finally, the formation control stability criterion and the corresponding numerical results are provided in Sect. 4.

2 Kinetic Model of the Overall Formulation System

2.1 Space Description of Formation System and Its Systematic

We consider a three-star system which is shown in Fig. 1, where $m_i (i = 1, 2, 3)$ represent all the members involved in the formation system, whose coordinates in inertial system (XOY) are nominated as r_i . We suppose that the spacecraft are limited in a certain plane, and the formation movement in configuration space can well be described by $x = (r_1, r_2, r_3)$. Formation systems usually maintain a fixed formation, and the whole formation can move by means of translation and rotation. Thus, the movement of the formations can be decomposed into three parts: deformation, translation, and rotation.

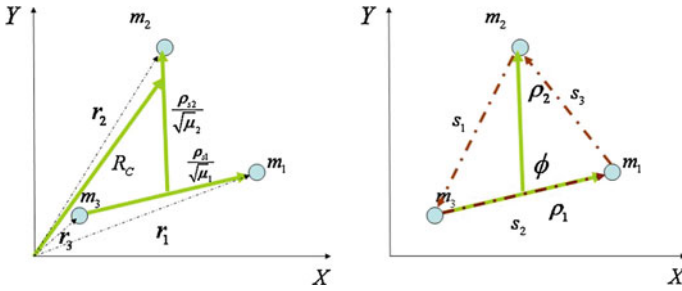


Fig. 1 The relationship between formation and coordinate

As is described in reference [2], we adopt the weighted Jacobi coordinates to eliminate the translation DoF. From Fig. 1, we can rearrange the coordinate transformation relationship as follows:

$$\begin{bmatrix} \rho_{s1} \\ \rho_{s2} \\ R_C \end{bmatrix} = \begin{bmatrix} \sqrt{\mu_1} & 0 & -\sqrt{\mu_1} \\ -\sqrt{\mu_2} \frac{m_1}{m_1+m_3} & \sqrt{\mu_2} & -\sqrt{\mu_2} \frac{m_3}{m_1+m_3} \\ \frac{m_1}{M} & \frac{m_2}{M} & \frac{m_3}{M} \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} \quad (1)$$

where $\frac{1}{\mu_1} = \frac{1}{m_1+m_3}$, $\frac{1}{\mu_2} = \frac{1}{m_1+m_3} + \frac{1}{m_2}$, $M = m_1 + m_2 + m_3$. $R_C \in \mathbb{R}^2$ represent the position of the formation centroid, and $(\rho_{s1}, \rho_{s2}) \in \mathbb{R}^2 \times \mathbb{R}^2$ denotes the shape and orientation of formation, respectively. In Robert and Matthias [2], a group of shape coordinates are defined as:

$$q^1 = \rho_1 = \|\rho_{s1}\|, q^2 = \rho_2 = \|\rho_{s2}\|, q^3 = \varphi = \arccos \frac{\rho_{s1} \times \rho_{s2}}{\|\rho_{s1}\| \|\rho_{s2}\|} \quad (2)$$

To simplify the design of controller, based upon the above coordinate, we also use the following shape coordinate:

$$s_2 = \frac{\rho_1}{\sqrt{\mu_1}} \quad (3)$$

$$s_1 = \left[\left(\frac{\rho_2}{\sqrt{\mu_2}} \sin(\varphi) \right)^2 + \left(\frac{\rho_2}{\sqrt{\mu_2}} \cos(\varphi) + \frac{m_1}{m_1+m_3} s_2 \right)^2 \right]^{\frac{1}{2}} \quad (4)$$

$$s_3 = \left[\left(\frac{\rho_2}{\sqrt{\mu_2}} \sin(\varphi) \right)^2 + \left(-\frac{\rho_2}{\sqrt{\mu_2}} \cos(\varphi) + \frac{m_3}{m_1+m_3} s_2 \right)^2 \right]^{\frac{1}{2}} \quad (5)$$

From Fig. 1, we can find that the above shape coordinate represents the distance between different members of the formation system, i.e., $s_1 = \|r_3 - r_2\|$, $s_2 = \|r_1 - r_3\|$, $s_3 = \|r_2 - r_3\|$. In fact, the shape coordinates $q = (\rho_1, \rho_2, \varphi) \in \mathcal{S}$ and $q = (s_1, s_2, s_3) \in \mathcal{S}$ are different coefficients of the same shape space \mathcal{S} , and

Eq. (2.3) reveals the transforming relationship between them. Thus, $\phi : \mathcal{S} \rightarrow \mathcal{S}, q = (\rho_1, \rho_2, \varphi) \mapsto \tilde{q} = (s_1, s_2, s_3)$.

2.2 Separation of System Energy

The separating process of system energy can be implemented by calculating the derivations of Eq. (2.1)

$$\begin{bmatrix} \dot{r}_1 \\ \dot{r}_2 \\ \dot{r}_3 \end{bmatrix} = \begin{bmatrix} \frac{m_3}{m_1+m_3} & -\frac{m_2}{M} & 1 \\ 0 & \frac{m_1+m_3}{M} & 1 \\ -\frac{m_1}{m_1+m_3} & -\frac{m_2}{M} & 1 \end{bmatrix} \begin{bmatrix} \dot{\rho}_{s1}/\sqrt{\mu_1} \\ \dot{\rho}_{s2}/\sqrt{\mu_2} \\ \dot{R}_C \end{bmatrix} \quad (6)$$

The total kinetic energy of the system is $K_0 = \frac{1}{2} \sum_1^3 m_i \|\dot{r}_i\|^2$. Substitute it into (2.1), we have:

$$K_0 = \frac{1}{2} \sum_1^3 m_i \|\dot{R}_C\|^2 + \frac{1}{2} \|\dot{\rho}_{s1}\|^2 + \frac{1}{2} \|\dot{\rho}_{s2}\|^2$$

where $K_C = \frac{1}{2} \sum_1^3 m_i \|\dot{R}_C\|^2$ denotes the translation kinetic energy of the formation centroid. $K = \frac{1}{2} \|\dot{\rho}_{s1}\|^2 + \frac{1}{2} \|\dot{\rho}_{s2}\|^2$ contains the deformation energy and the overall rotational kinetic energy, and

$$K = \frac{1}{2} (\omega + A\dot{q})^T I_l (\omega + A\dot{q}) + \frac{1}{2} \dot{q}^T M \dot{q} \quad (7)$$

where $\frac{1}{2} (\omega + A\dot{q})^T I_l (\omega + A\dot{q})$ denotes the overall rotational kinetic energy, and $\frac{1}{2} \dot{q}^T M \dot{q}$ is the inner deformation kinetic energy, ω is the most commonly used angular velocity, A reflects the couple matrix of interior deformation and overall rotational movement, I_{locked} is the corresponding rotational inertia moment when the formation is locked. M is the generalized mass in the shape space \mathcal{S} .

2.3 Formation Dynamics

When considering the interaction among various members of the formation, we can rearrange the Lagrange function as:

$$L = \frac{1}{2} \sum_1^3 m_i \|\dot{R}_C\|^2 + \frac{1}{2} (\omega + A\dot{q})^T I_{locked} (\omega + A\dot{q}) + \frac{1}{2} \dot{q}^T M \dot{q} - V \quad (8)$$

Together with the adopted model in Sect. 2.1, we assume that $\omega = (0, 0, \dot{\theta})^T$, $\dot{q} = (\dot{\rho}_1, \dot{\rho}_2, \dot{\varphi})$. According to the Lagrange equation $\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}^i} \right) - \frac{\partial L}{\partial x^i} = Q_i$, we have:

$$M\ddot{R}_C = Q_C, \quad (9)$$

$$(\rho_1^2 + \rho_2^2)\ddot{\theta} + \rho_2^2\ddot{\varphi} + \frac{\partial V}{\partial \theta} = Q_\theta, \quad (10)$$

$$\ddot{\rho}_1 - \rho_1\dot{\theta}^2 + \frac{\partial V}{\partial \rho_1} = Q_{\rho_2}, \quad \rho_2^2\ddot{\varphi} + \rho_2^2\ddot{\theta} + \frac{\partial V}{\partial \varphi} = Q_\varphi \quad (11)$$

where Q_i is the nonconservative control input. Equations (2.9) and (2.10) represent the overall rigid translational and rotational kinetic equations of this formation system. Equation (2.11) gives the kinetic equations of the deformation. In particular, when $Q_C = Q_\theta = \frac{\partial V}{\partial \theta} = 0$, the translation and rotation symmetry are guaranteed.

$$p_C = \frac{\partial L}{\partial \dot{R}_C} = M\dot{R}_C \quad (12)$$

$$p_\theta = \frac{\partial L}{\partial \dot{\theta}} = (\rho_1^2 + \rho_2^2)\dot{\theta} + \rho_2^2\dot{\varphi}. \quad (13)$$

3 Stability and Formation Control

In this section, we use the concept of relative balance [3] to study the formation stability problem. Suppose that the corresponding Hamilton function of the formation which is illustrated in Eq. (2.7) or (2.8) is H and the canonical coordinates are $z = (q, p) \in \mathcal{P}$, then the Hamilton form of the kinematic system is $\frac{d}{dt}z = J\nabla H$, where J is the symmetrical matrix [8]. Before discussing the formation stability problem, we first provide some related definitions and lemmas.

Definition 1 (*Relative balance*) [3]: We say that the canonical coordinate $z_e = (q_e, p_e) \in \mathcal{P}$ is relative balance if $X_H(z_e) \in T_{z_e}(G \times z_e)$, i.e., the value of Hamilton vector field in z_e is of the tangent space $T_{z_e}(G \times z_e)$ that comes from the symmetry group of trajectory $G \times z_e$.

Lemma 1 (Criterion of relative balance): *If the Hamilton function has the same form as the summation of kinetic and potential energy, we can set the generalized potential energy V_μ as the summation of natural potential and centrifugal potential: $V_\mu = V + \frac{1}{2}(\omega + A\dot{q})^T I_l(\omega + A\dot{q})$. Thus, we say $z_e = (q_e, p_e)$ is relative balance if and only if it satisfies $\frac{\partial}{\partial q} V_\mu(q_e) = 0, p_e = \frac{\partial L(q_e, \dot{q}_e)}{\partial \dot{q}}$.*

Lemma 2 (Stability of relative balance) [3]: *Under the same condition as lemma 1, if we can further guarantee that $\frac{\partial^2 V_\mu(q_e)}{\partial q^2} > 0$, then the relative balance $z_e = (q_e, p_e)$ is stable in the relative space \mathcal{P} and is Lyapunov stable in the reduced space \mathcal{P}/G .*

3.1 Formation Control with Zero Angular Momentum

We first consider the case where $Q_\theta = \frac{\partial V}{\partial \theta} = 0$, i.e., the system has rotational symmetry. In particular, when $p_\theta = \frac{\partial L}{\partial \dot{\theta}} = (\rho_1^2 + \rho_2^2)\dot{\theta} + \rho_2^2\dot{\phi} = 0$, we use the potential fields which exist between different members $V_1(s_1), V_2(s_2), V_3(s_3)$ to describe the input of deformation control. Now, the transformation between two groups of shape coordinates can be written as $\phi : S \rightarrow S, q = (\rho_1, \rho_2, \varphi) \mapsto \tilde{q} = (s_1, s_2, s_3)$. Suppose that $V(\tilde{q}) = V_1(s_1) + V_2(s_2) + V_3(s_3)$ and $V_i(s_i)$ is lower bounded, which indicates $C_i < -\infty$ and $V_i(s_i) > C_i$, then we have following conclusions:

Conclusion 1 When $p_\theta = 0$,

- (1) If $\frac{\partial V(\tilde{q}_e)}{\partial s_i} = 0$, then $q_e = \phi^{-1}(\tilde{q}_e) \in S$ or $\tilde{q}_e \in S$ is a balanced formation;
- (2) If $\frac{\partial V(\tilde{q}_e)}{\partial s_i} = 0$ and $\frac{\partial^2 V(\tilde{q}_e)}{\partial s_i \partial s_j} > 0$, then $q_e = \phi^{-1}(\tilde{q}_e) \in S$ or $\tilde{q}_e \in S$ is a balanced formation. Furthermore, if there exists damping force $Q_i(s, \dot{s})$, which equally means that $Q_i(s, 0) = 0, \dot{s}_i Q_i(s, \dot{s}) \leq 0$, then the formation is asymptotically stable;
- (3) Based upon condition (2), there exist inputs of formation control $\tilde{u}_{D1} = -\nabla V(\tilde{q})$ and $\tilde{u}_{D2} = -\nabla V(\tilde{q}) + Q(s, \dot{s})$ such that the system stability and asymptotical stability is guaranteed, and the corresponding formations are $\tilde{q}_e \in S$ or $q_e = \phi^{-1}(\tilde{q}_e) \in S$.

With regard to conclusion (1), if $\frac{\partial V(\tilde{q}_e)}{\partial s_i} = 0$, then we have

$$\frac{\partial (V \circ \phi)(q_e)}{\partial q_i} = \frac{\partial V(\phi(q_e))}{\partial \tilde{q}_j} \frac{\partial \tilde{q}_j}{\partial q_i} = \frac{\partial V(\tilde{q}_e)}{\partial \tilde{q}_j} \frac{\partial \tilde{q}_j}{\partial q_i}(q_e) = 0 \tag{14}$$

From lemma 1, we know that $q_e = \phi^{-1}(\tilde{q}_e) \in S$ is a balanced formation.

For conclusion (2), we have

$$\frac{\partial}{\partial q_j} \left(\frac{\partial(V \circ \phi)}{\partial q_i} \right) = \frac{\partial}{\partial q_j} \left(\frac{\partial V(\tilde{q})}{\partial \tilde{q}_r} \frac{\partial \tilde{q}_r}{\partial q_i} \right) = \frac{\partial \tilde{q}_r}{\partial q_i} \frac{\partial^2 V(\tilde{q})}{\partial \tilde{q}_r \partial \tilde{q}_s} \frac{\partial \tilde{q}_s}{\partial q_j} + \frac{\partial V(\tilde{q})}{\partial \tilde{q}_r} \frac{\partial^2 \tilde{q}_r}{\partial q_i \partial q_j} \quad (15)$$

In the balanced point, we have $\frac{\partial V(\tilde{q}_e)}{\partial \tilde{q}_r} = 0$. In addition, it can be easily verified that $\frac{\partial^2(V \circ \phi)(q_e)}{\partial q_i \partial q_j} > 0$. Here, we denote $\tilde{V} = V(\phi(q)) - C \geq 0$ as the potential energy of the formation. From lemma 2, we know that the balanced formation $q_e = \phi^{-1}(\tilde{q}_e) \in S$ is stable. Furthermore, if the complete damping is taken into account, the asymptotical stability is also guaranteed.

With regard to conclusion (3), the controlled formation is seen as a dynamical system within which various potential fields interact with each other, and the discussion of conclusion (2) also holds true.

Remark 1 It is worth noting that the conclusion above is similar to the Lagrange theorem in the stability theory [8, 12, 13] of dynamical systems. The difference is that balanced formation we discussed here is a fixed point in shape space rather than the fixed point in configuration space in the general sense. Note that the property of the singular point of the function is independent of the choice of coordinates, then conclusion (1) and conclusion (2) are natural corollary of Lemma 1 and Lemma 2.

3.2 Formation Control with Nonzero Angular Momentum

In this section, we further study the case where $Q_\theta = \frac{\partial V}{\partial \theta} = 0$. Suppose that the conserved quantity in (2.12) is not zero, i.e., $p_\theta = \frac{\partial L}{\partial \dot{\theta}} = (\rho_1^2 + \rho_2^2)\dot{\theta} + \rho_2^2\dot{\phi} = \mu \neq 0$, which means that the whole system has rigid rotation when the formation motion is in the steady state.

Let $V(\tilde{q}) = V_1(s_1) + V_2(s_2) + V_3(s_3)$ denote the interactions among various formation members. The inertia tensor of the system is $I_l(q)$, and $\mu \neq 0$. By choosing $p = [0, 0, \mu]^T$ and the valid potential $V_\mu(q) = V(\phi(q)) + \frac{1}{2}p^T I_l^{-1}(q)p$, combining with Lemmas 1 and 2, we have the following conclusions:

Conclusion 2 $p_\theta = \mu \neq 0$

- (1) If $\frac{\partial V_\mu(q_e)}{\partial q_i} = 0$, then $q_e \in S$ is a balanced formation.
- (2) If $\frac{\partial V_\mu(q_e)}{\partial q_i} = 0$ and $\frac{\partial^2 V_\mu(q_e)}{\partial q_i \partial q_j} > 0$ (positive definite), then $q_e \in S$ is a stable formation.

Conclusions (1) and (2) are direct applications of Lemmas 1 and 2. Thus, we omit the proof here.

3.3 Coordinate Transformation of Control Law

The transformation from the shape space to the inertial coordinate system is needed when the control law is implemented. The transformation relationship at the situation discussed in Sect. 3.1 will be given below. Notice that the vector itself is independent of the coordinate system, then $\nabla_{\tilde{q}}V(\tilde{q}) = \nabla_rV(\psi(r))$, where $\psi : (r_1, r_2, r_3) \mapsto (s_1, s_2, s_3, \theta, R_C)$ represents the transformation of the coordinates. Thus, the potential field function of the i^{th} spacecraft is:

$$u_i = -\frac{\partial}{\partial r_i}V(\psi(r)) = -\frac{\partial V(\psi(r))}{\partial \tilde{q}_j} \cdot \frac{\partial \tilde{q}_j}{\partial r_i} \quad (16)$$

Since $Q_i(s, 0) = 0$, $\dot{s}_i Q_i(s, \dot{s}) \leq 0$, we can set

$$Q_i(s, \dot{s}) = -\frac{\partial R(\dot{s}, s)}{\partial \dot{s}_i} \quad (17)$$

where $R(\dot{s}, s)$ is Rayleigh dissipation function:

$$R(\dot{s}, s) = \frac{1}{2} \sum_{i=1}^3 k_i \dot{s}_i^2, \quad k_i > 0 \quad (18)$$

Note that m_1 interacts with m_3 and m_2 through s_2 and s_3 , respectively, then the potential field $V_2(s_2)$, $V_3(s_3)$ act on m_1 , and the input on m_1 is

$$u_1 = -\nabla V_2(s_2) - \nabla V_3(s_3) \quad (19)$$

which can be further written as

$$u_{1x} = -\frac{\partial V_2(s_2)}{\partial s_2} \frac{(x_1 - x_3)}{s_2} + \frac{\partial V_3(s_3)}{\partial s_3} \frac{(x_2 - x_1)}{s_3} \quad (20)$$

$$u_{1y} = -\frac{\partial V_2(s_2)}{\partial s_2} \frac{(y_1 - y_3)}{s_2} + \frac{\partial V_3(s_3)}{\partial s_3} \frac{(y_2 - y_1)}{s_3} \quad (21)$$

In addition, the damping control input on m_1 can be derived from Eq. (2.13)

$$(u_{1x}^d, u_{1y}^d) = Q_2 - Q_3 = \frac{1}{s_2} (-k_2 \dot{s}_2)(x_1 - x_3, y_1 - y_3) + \frac{1}{s_3} (k_3 \dot{s}_3)(x_2 - x_1, y_2 - y_1) \quad (22)$$

where $s_2 = \sqrt{(x_1 - x_3)^2 + (y_1 - y_3)^2}$, $s_3 = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$.

3.4 Potential Field Selection in Formation Control

The aforementioned analysis indicates that the control of deformation in formation control depends on the interacting potential field $V(\tilde{q}) = V_1(s_1) + V_2(s_2) + V_3(s_3)$. Here, we adopt the potential field with linear spring form

$$V(\tilde{q}) = \frac{1}{2} \sum_{i=1}^3 k_i (s_i - d_i)^2, \quad k_i > 0 \tag{23}$$

4 Numerical Results

In this section, we verify the effectiveness of the proposed control laws through numerical simulation. We still consider a three-spacecraft system with $m_i = 1, \quad i = 1, 2, 3$. They are initially located on a circumference whose radius is r_0 , and the phases are chosen as $\frac{\pi}{6}, \frac{\pi}{2}, \frac{3\pi}{2}$, respectively. The radius of the desired formation is $r_d = 1$, and the mutual distances are set to be the same. Then, the corresponding coordinates of the desired formation should be $s_{di} = \sqrt{3} \approx 1.7321, \quad i = 1, 2, 3, k = [10, 10, 10], k_d = [1, 1, 1]$.

In Fig. 2, each member of the formation system distributes in a circumference whose radius is 2 without initial velocity. And the mutual distance between them is equal with $s_{di} = \sqrt{3} \approx 1.7321$. Figure 2a shows that the position of each member in the inertial coordinate system, while Fig. 2b shows the time history of the mutual distance among members of the formation in the shape space, i.e., the change of the shape coordinates.

The physical process presented in Fig. 3 is similar as that in Fig. 2, while the main difference between them is that the formation system having an initial

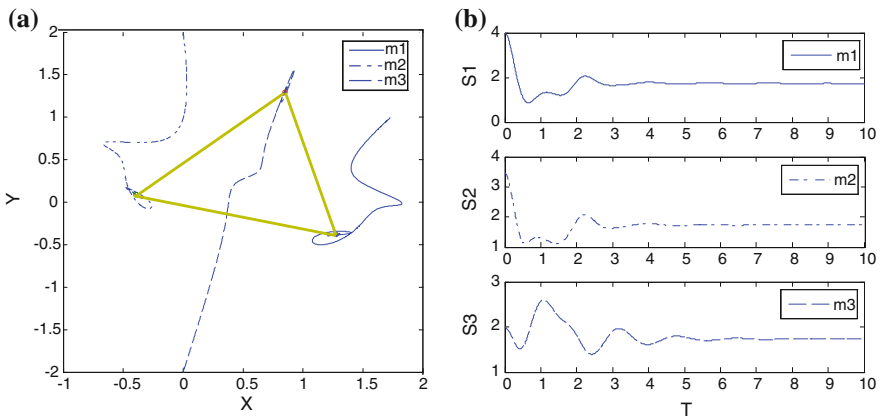


Fig. 2 (a) $p_0 = 0, r_0 = 2$ (b) $p_0 = 0, r_0 = 2$

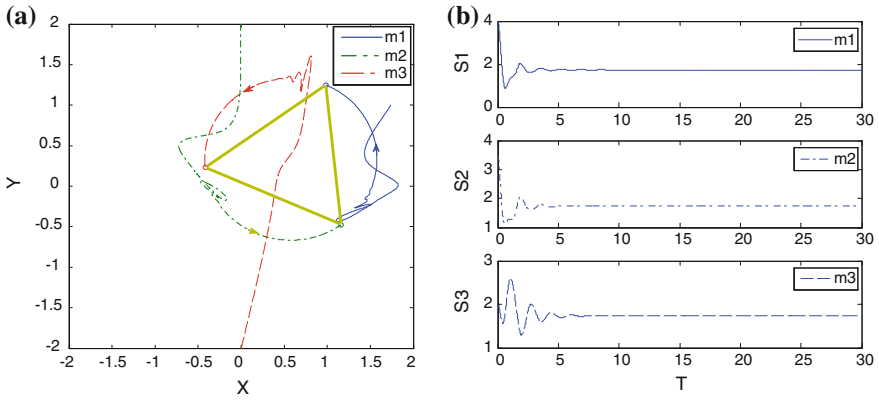


Fig. 3 (a) $p_\theta \neq 0$ (b) $p_\theta \neq 0$

velocity $\dot{r}_0 = [0.2, 0, 0, -0.2, 0, 0]$, and $p_c = 0, p_\theta \neq 0$. It can be observed from Fig. 3 that rigid rotation exists when the formation system is in steady state, and the balanced formation corresponds to the extreme value points of the generalized potential, which means that the overall rigid rotation renders the balance point of the system drift.

Figure 4 shows the change of formation in the inertial space. Here, the linear momentum of the formations system is $p_c \neq 0$, and the initial condition is $r_0 = 2, \dot{r}_0 = [0.5, 0.4, 0.5, 0.4, 0.5, 0.4]$. Figure 4a shows the rigid translational movement of system after the steady state. Figure 4b shows the formation change when $p_c \neq 0, p_\theta \neq 0$, and the initial condition is $r_0 = 2, \dot{r}_0 = [0.5, 0, 0, 0.4, 0, 0.4]$. It can be seen from the simulation results that there exist both rigid translational and rotational movement when the formation enter into the steady stage.

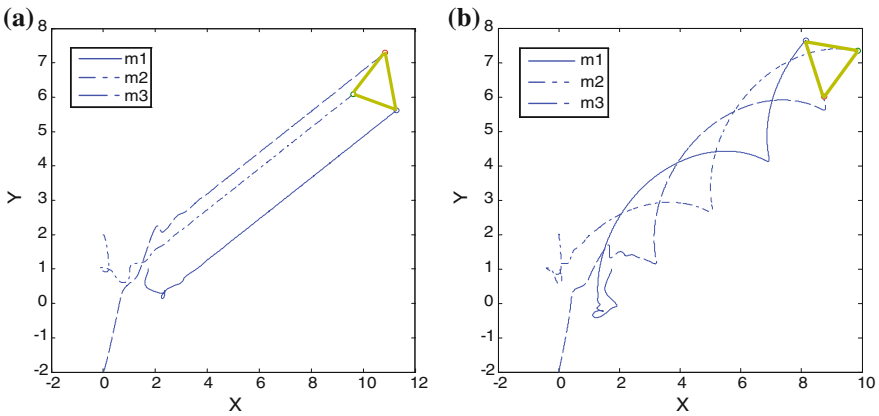


Fig. 4 (a) $p_\theta = 0, p_c \neq 0, r_0 = 2, \dot{r}_0 = [0.5, 0.4, 0.5, 0.4, 0.5, 0.4]$ (b) $p_\theta \neq 0, p_c \neq 0, r_0 = 2, \dot{r}_0 = [0.5, 0, 0, 0.4, 0, 0.4]$

5 Conclusions

This paper studies the formation control of spacecraft from the perspectives of energy. The formation system was viewed as a dynamic system with coupling characteristics. Based upon shape coordinates description, the system dynamic equations with both internal and external separation forms are derived. The formation is decomposed into deformation, rigid translation, and rigid rotation. In addition, the mathematical definition of the balanced formation and formation stability are provided, and the stability criterion of the formation control is also presented. The formation control laws are designed based on the energy concept, and the corresponding stability issue is also studied. Numerical simulations are conducted to demonstrate the effectiveness of the proposed approach.

Acknowledgments This work was financially supported by the Tsinghua Self-innovation Project (Grant No: 20111081111) and the National Natural Science Foundation of China (Grant Nos: 2009CB724000, 2012CB821206, 61202332, 61203354).

References

1. Daniel PS, Fred YH, Scott RP (2004) A survey of spacecraft formation flying guidance and control (Part II). In: Proceeding of the 2004 American control conference. Massachusetts, Boston
2. Robert GL, Matthias R (1997) Gauge fields in the separation of rotations and internal motions in the n-body problem. *Rev Mod Phys* 69(1)
3. Jerrod EM (1992) Lectures on mechanics. Cambridge University Press, New York
4. Naomi EL, Edward F (2001) Virtual, artificial potentials and coordinated control of groups. In: Proceeding 40th IEEE conference on decision and control
5. Zhang F (204) Geometric cooperative control of formations, PhD Thesis, University of Maryland, US
6. Edward AF (2005) Cooperative vehicle control, Feature tracking and ocean sampling, PhD Thesis, Princeton University
7. Anthony MB, Dong EC, Naomi EL, Jerrold EM (2001) Controlled Lagrangians and the stabilization of mechanical systems. II. Potential shaping. *IEEE Trans Autom Control* 46 (10):2253–2270
8. Malta G (2006) Theoretical mechanics. Higher Education Press, Beijing
9. Willems JC (1991) Paradigms and puzzles in the theory of dynamical systems. *IEEE Trans Autom Control* 36(3):259–294
10. Ortega R, van-der-Schaft AJ, Mareels I, Maschke B (2001) Putting energy back in control. *IEEE Control Syst Mag* 21(2):18–33
11. Slotine JJ (1988) Putting physics in control—The example of robotics. *IEEE Control Syst Mag* 8(6):12–18
12. Wang Z (1992) Stability of motion and its application. High Education Press, Beijing
13. Lin H (2001) Stability theory. Peking University Press, Beijing

An Intelligent Conflict Resolution Algorithm of Multiple Airplanes

Jingjuan Zhang, Jun Wu and Rong Zhao

Abstract To resolve the conflict of multiple airplanes in free flight, a genetic algorithm which can plan the routes quickly and accurately is proposed, and a simulation platform using the powerful MATLAB is built. The experimental results of flight conflict resolutions of 2, 3, and 5 airplanes, especially 10 similar airplanes approaching one another in a symmetrical manner, in the round field 300 km in diameter, show that the algorithm can solve the conflicts effectively within 300 s. The new route line is smooth, and airplanes can straight back to the intended routes after solving the conflicts. The algorithm matches the requirements of feasibility, rapidity, safety, fuel efficiency, and passenger comfort in the real flight operations. This work provides theoretic and design references for the development of the safety technology in the aviation's next-generation global CNS/ATM system.

Keywords Free flight · CNS/ATM · Conflict resolution · Genetic algorithm

1 Introduction

Aviation transportation, playing a key role in the activities of the modern world economy, is one of the industries with the fastest growth in the world economy. The air transport systems on a worldwide scale have been increasing steadily for

J. Zhang (✉) · J. Wu
Beihang University, Beijing 100191, China
e-mail: zhangjingjuan@buaa.edu.cn

J. Wu
e-mail: dugujian5@sina.com

R. Zhao
AVIC Avuonics Co. Ltd, Beijing 100191, China
e-mail: honorzhao@163.com

the last 20 years. In order to cope with the continuous growth of traffic demand, changes over the air traffic system are being ruled by the so-called CNS/ATM (Communication, Navigation, Surveillance/Air Traffic Management) [1, 2] paradigm. With the development of data link between airplanes and ground-based control centers, GPS, and 4D Flight Management Systems [3], the CNS/ATM will transform the current air traffic system, based on voice communications and independent (radar) surveillance, into a digital network distributed in large scale.

Under proposed air traffic management concepts such as free flight, aircraft would have more flexibility to follow efficient routes in response to changing conditions. However, they bring up new challenges to be solved. The loss of an airway structure may make the process of resolving conflicts between aircrafts more complex. Accordingly, the aircraft conflict resolution is the key technology whether the free flight will be realized and has been the object of intensive research [4, 5]. Under a popular technology named “Automatic Dependent Surveillance-Broadcasting” (ADS-B) [6, 7], it is possible that the whole airspace situation can be seen by every pilot and air traffic controller to assess risks.

The goal of the work presented here was to resolve the conflict between 2 airplanes or among 3 ones, even multi-airplanes quickly, accurately, and humanely. We concentrated on the modeling of the problem and the realistic requirements. Considering genetic algorithm (GA) is a global optimal searching algorithm based on the biological evolutionism [8]. An improved conflict resolution algorithm based upon GA was proposed. Then, we used the powerful programming software MATLAB to design the algorithm simulation.

2 Model Construction

In a real flight, the problem needs to be appropriately simplified in order to improve the speed of the algorithm because it is extremely important to supply pilots the optional air route that would not meet the confliction as quickly as we can.

Airplanes, especially the civilian aircrafts, normally fly on fixed heights during their flight routes except while taking off and landing [2]. On the other hand, taking into account the comfort of passengers and fuel consumption, the pilots normally do not change the vertical heights of airplane, but adjust the heading when avoiding conflicts in flight routes except while taking off and landing [9]. Therefore, the three-dimensional problem can be simplified into two-dimensional problem. In addition, the flight speed and height are assumed to be unchangeable for the simplicity of the problem. When avoiding collision, the maximization of saving time and fuel is realized by selecting the shortest routes.

Based on these simplifications of the problem, the confliction zone is set to a round field 300 km in diameter. There are n airplanes, and they fly in a straight line across the center point of the zone. They simultaneously fly in the zone, and the angle between any two neighboring predetermined airlines is $180^\circ/n$. Thus, all the

airplanes will collide in the center point of the zone. That will be the worst case that we focus on.

Most related researches [10–12] on conflict resolution based on GA assumed that there are only three direction choices for airline change, including the original flight direction, 30° left or 30° right. Their results were like the polygonal line that the real airlines must not be. What is more, all aircrafts entered and left the sector at the same time. The exit point was the point that would be reached at exit time if each airplane was able to fly straight to it without any heading alteration. This assumption was a little simple and unpractical.

In this paper, the model has been improved. The realistic conditions that civilian airplanes do not turn in big angles suddenly and airlines are smooth are considered, and we assume that the pilots could choose arbitrary angle between a certain angle (such as 30°) left-inclined and a certain angle (such as 30°) right-inclined. Therefore, the coding method can be determined.

When an airplane enters the zone at the entry point, its position and its current heading are accurately known, taking advantage of the information discovered by ADS-B. According to the safety regulations of air traffic management in China, any two airplanes are in conflict if, at any time, they are closer than 20 km. The constraints is given by

$$d = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} > 20 \quad (1)$$

where (x_i, y_i) and (x_j, y_j) are the coordinates of airplane i and airplane j in the horizontal airplane.

Therefore, we choose 10 km as one step the airplanes move in order to raise the conflict detection accuracy. The 10 km airline is no longer seen as a point but as an airline segment. All routes in the initial population are of fixed length, 300 km multiplied by a length coefficient. The direction of the next airline segment is an arbitrary angle between 30° left and 30° right, taking the one of the current airline segment for reference. The rest can be done in the same manner. Each trajectory data were encoded with the coordinates of the points.

The airplanes could not have straight airlines when avoiding conflict, but should have the shortest ones under the condition of no conflicts. The objective function in the physical space is given by

$$y = \min \sum_{i=1}^n S_i \quad (2)$$

where n is the number of aircraft within the zone, S_i is the length of the airplane, and i is the route of the zone.

If there is no conflict, the distance d_i from the entrance point to the exit point is computed for each airplane i . The distance d_i is then given by

$$d_i = \sum_{i=1}^{m-1} \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} \quad (3)$$

where (x_i, y_i) and (x_{i+1}, y_{i+1}) are the current step and the next step's coordinates of an airplane and m is the number of steps within the zone.

The fitness f is given by

$$f = \sum_{i=1}^n d_i/t \quad (4)$$

where n is the number of the airplanes in the zone and t is the coefficient to adjust the range of f . The system checks whether there is a conflict between two (or more planes) according to the trajectories calculated (i.e., it checks whether two airplanes are ever closer than 20 km). If a conflict is found, then the fitness of the sequence is then given by

$$f = \sum_{i=1}^n d_i/t + d/c \quad (5)$$

where d is the distance of the two airplanes calculated and c is the coefficient to adjust the range of d .

The values of t and c must be correctly chosen to guarantee a correct behavior of the function: If t is too large, the value of d_i/t will be closed to 0, such routes that there were on conflict will have the larger fitness, and if c is too large, the value of d/c will be closed to 0, such routes that guaranteed the shortest distances but contained the conflict will have the larger fitness.

3 Genetic Algorithm and Implementation

In order to avoid plane conflicts in free flight, GAs are adopted in flight route selection. GAs are probabilistic search algorithms for global optimization [13–15], which simulate the biological genetic and evolutionary process in the natural environment. GAs start by initializing a set (population) containing a selection of encoded points of the search space (individuals). By decoding the individual and determining its cost, the fitness of an individual can be determined, which is used to distinguish between better and worse individuals. A GA iteratively tries to improve the average fitness of a population by construction of new populations. A new population consists of individual (children) constructed from individuals of the old population (parents) by the use of recombination operators. Better (above average) individuals have higher probability to be selected for recombination than other individuals (survival of the fittest). After some criterion is met, the algorithm returns the best individuals of the population.

1. The program calculates the fitness of each individual, ranks them in the fitness descending order, and then eliminates the back half of the list.
2. The individuals are randomly paired, and the locations of the intersections are randomly selected. It is specific that the two parts in each pair after the intersection are compromised to replace the original ones. This is one of the improvements in the program.
3. Some individuals are randomly chosen based on a certain probability and then mutate the heading angle between 30° left and 30° right in a random step.

4 Simulation

In this paper, MATLAB was applied to realize the above-mentioned improved GA, simulates the flying process of airplanes, and records the experimental data.

MATLAB language is an array programming language, which is very suitable for calculating in the algorithm. What is more, MATLAB has a complete graphics capability and can realize the visualization of calculation results.

As the individuals are encoded by vectors, it is a good choice to use MATLAB to solve this problem. In addition, it is intuitive and convenient to visualize the results.

5 Presentation of Results

We solved the problem with two, three, and five planes which are in conflict at the center of the zone. Six hundred initial random sequences were generated. The crossing parameter was set to 70 % and the mutation parameter to 2 %. Yet, the length coefficient is correspondingly changed with the number of airplanes. Three versions of programs were separately designed for the conflict resolution of two, three, and five planes.

First of all, the program solved the 2 airplanes conflict with the scene that one airplane is flying from west to east and the other from south to north.

Figure 1 showed the convergence process of the improved GA, and the subfigures a, b, and c show the results when evolving to the initial, the 201st generation, the 324th generation. From the figure, we can see that the new air routes made the interval of two airplanes always more than 20 km.

The problem of the 3 planes being in conflict at the same point (a plane crossing the sector from the lower left corner to the upper right corner was added to the previous example) was also solved. The start headings of adjacent planes were 60° apart. The results were computed with the length coefficient increasing in some sort. In order to see the convergence process of the improved GA, the route planning process was recorded, as shown in the Fig. 2. The subfigures a, b, and c

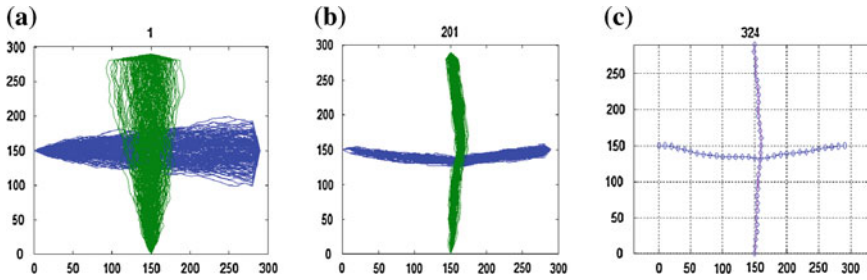


Fig. 1 The process diagrams of 2 airplanes procedure. **a** The initial. **b** The 201st generation. **c** The 324th generation

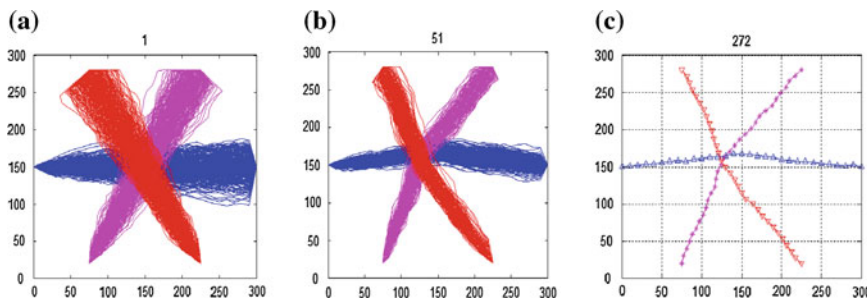


Fig. 2 The process diagrams of 3 airplanes procedure. **a** The initial. **b** The 51st generation. **c** The 272nd generation

show the results when evolving to the initial, the 51st generation, and the 272nd generation.

Two planes crossing the sector from the lower left corner to the upper right corner were added to the 3 planes example. The start headings of adjacent planes were 36° apart. Of course, the length coefficient increased a little more.

The convergence process with 5 planes was recorded in the Fig. 3.

Additionally, in order to study the conflict avoidance strategy, a hypothetical multiple conflict involving ten similar airplanes approaching one another in a symmetrical manner is discussed.

The convergence process with 10 planes was recorded in the Fig. 4.

The “-” line showed routes of planes crossing the sector from west to east and from the lower corner to the upper corner. The “-△” line showed routes of planes crossing the sector from east to west and from the upper corner to the lower corner. There is a common phenomenon that the directions of the airplanes heading changing are either all in the airplanes’ left or all in the right. That is the conflict avoidance strategy.

Additionally, the average situations of 10 random sample experiences of 2, 3, and 5 planes problem were shown in Table 1.

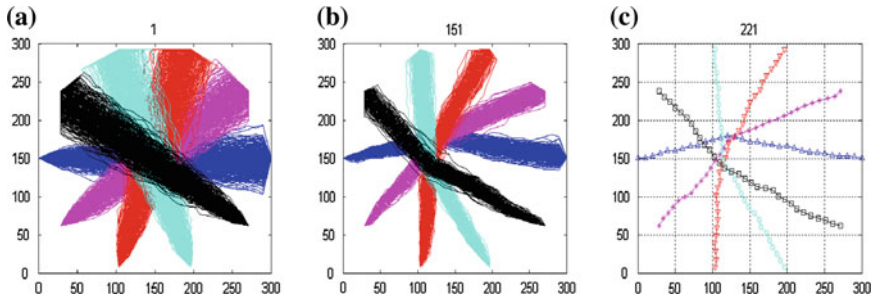


Fig. 3 The process diagrams of 5 airplanes procedure. **a** The initial. **b** The 151st generation. **c** The 329th generation

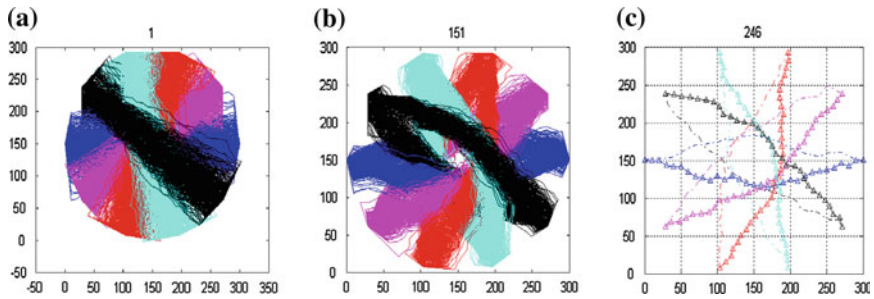


Fig. 4 The process diagrams of 10 airplanes procedure. **a** The initial. **b** The 151st generation. **c** The 212th generation

Table 1 The 10 random sample results of conflict resolution

The number of airplanes	Average evolutionary generations	Average convergence time (s)
2	228	60.8
3	290	121.9
5	334	290.3

In order to shorten the time to calculate and enhance the convergence of algorithm, such methods can be taken: reduce the number of individuals in the population and shorten the code length. In fact, reducing the number of individuals in the population can shorten the iteration time, but reduces the population diversity which makes it hard to find the best result. However, shortening the code length will reduce the airplane’s minimal yaw angle, but it has little effect on the final result, taking into account of the actual minimal yaw angle.

The average situations of 10 random sample experiences of 2, 3, and 5 planes problem after shortening the code length were shown in Table 2.

Table 2 The 10 random sample results of conflict resolution

The number of airplanes	Average evolutionary generations	Average convergence time (s)
2	200	56.3
3	269	96.9
5	228	190.1

By comparing Tables 1 and 2, it is obvious that shortening the code length is an effective way to enhance the convergence of algorithm and shorten the time to calculate.

6 Conclusions

1. An improved GA is developed to resolve flight conflicts in this paper, and the MATLAB is used to design the algorithm simulation.
2. The experimental results of 2, 3, and 5 airplanes all met the requirement in a real flight such as safety, rapidity, feasibility, and so on. The predictable confliction has been solved in a short time so that pilots could have enough time to do the emergency maneuver. What is more, the routes were so smooth when heading angle was adjusted that met the practical requirements of pilots and passengers.
3. The conflict avoidance strategy is studied. The conflict resolution results show that the directions of the airplanes heading changing are either all in the airplanes' left or all in the right.
4. A method to shorten the iteration time is proved to be feasible. It takes less time to solve the predictable confliction, and the route is smooth and optimized.

This work provides theoretic and design references for the development of the safety technology in the aviation's next-generation global CNS/ATM system.

References

1. Crow R (2002) Aviation's next generation global CNS/ATM system. IEEE Pos Location Navig Symp 291–298
2. Jun Zhang (2003) Modern air traffic management. Beihang University Press, Beijing [in Chinese]
3. Ahlstrom K, Torim J (2002) Future architecture of flight control systems. IEEE AESS Syst Mag 12:21–27
4. Profit R (1995) Systematic safety management in the air traffic services. Euromoney Publications, London
5. Brooker P (2002) Future air traffic management: quantitative enroute safety assessment. J Navig 55:197–211

6. EUROCAE (2006) Safety, performance and interoperability requirements document for ADS-B NRA application. ED-126
7. ICAO (2003) ADS-B study and implementation task force. In: CNS/MET-ATM (Working Paper 8), Bangkok
8. Goldberg D (1989) Genetic Algorithms in Search, Optimization and Machine Learning. Addison Wesley, Reading, MA
9. Kuchar JK, Yang LC (2000) A review of conflict detection and resolution modeling methods. IEEE Trans Intell Trans Syst 1(4):179–189
10. Alliot JM, Gruber H, Joly G et al (1993) Genetic algorithms for solving air traffic control conflicts. In: The ninth conference on artificial intelligence for applications
11. Liu X, Hu M, Xiangning D. Application of genetic algorithms for solving flight conflicts. J Nanjing Univ Aeronaut Astronaut 34(1). (in Chinese)
12. Yang S, Dai F (2007) Conflict resolution in free flight based on an immune genetic algorithm. Aeronaut Comput Tech 37(1). (in Chinese)
13. Lawrence D (1990) Genetic algorithms and simulated annealing. Morgan Kaufman Publishers, Inc., Los Altos
14. Alaeddini A (2008) Efficient webs for conflict resolving maneuvers based on genetic algorithms, master of science dissertation. Sharif University of Technology, Tehran
15. Gerdes IS (1994) Application of genetic algorithms to the problem of free-routing for aircraft. In: IEEE, pp 536–541

A Novel End-to-End Authentication Protocol for Satellite Mobile Communication Networks

Xiaoliang Zhang, Heyu Liu, Yong Lu and Fuchun Sun

Abstract This paper establishes a model of satellite mobile communication network end-to-end authentication and designs a distributed end-to-end authentication protocol for the model including initial authentication, authentication status inquire, and re-authentication. It also analyzes the safety of the protocol by the formal verification and makes performance analysis by simulation experiments.

Keywords Satellite mobile communication · End-to-end · Authentication · Protocol

1 Introduction

There are three types of end-to-end authentication protocols: private-key-based authentication mechanism with the reliable third party, for example, Kerberos System of MIT [1–2]; private-key-based authentication mechanism without the reliable third party, for example, the authentication program proposed by Chang

X. Zhang (✉)

National Key Laboratory of Integrated Information System Technology,
Chinese Academy of Sciences, Beijing 100190, China
e-mail: zhangxl9497@gmail.com

H. Liu · Y. Lu · F. Sun

State Key Laboratory of Intelligence Technology and Systems, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
e-mail: liuheykicker@yahoo.com.cn

Y. Lu

e-mail: lysky007@126.com

F. Sun

e-mail: fcsun@mails.tsinghua.edu.cn

et al. [3]; public-key-based authentication mechanism without the third party, for example, X. 509 [4] and Diffie-Hellman switching protocol [5].

Under the circumstance of satellite network, due to the long latency of data transferring via satellite links, authentication protocols with the third party are not likely to be adopted. Therefore, albeit the computational cost is high for the end-to-end protocols based on public-key systems, it is the small number of mutual times that makes it acquire better comprehensive performance, for example, authentication protocols proposed by Xu [6] etc. and Tu [7] etc.

The features of satellite network communications like long latency, intensive mobility, high bit error rates (BERs) restrict the communicational efficiency. To solve this problem, this paper tries to make use of the distribute authentication management ability of gateways and the confidential relationships between them to propose a method of effectively decreasing space data transferring affect during mutual process of the protocol, which largely improve the efficiency of end-to-end authentication.

2 Authentication Model

This part proposes an authentication model for end-to-end communication in mobile satellite network. Then, the functions of every network element are described and communicational reference nodes are precisely defined. Finally, we briefly introduced the authentication information involved in the authentication process (Fig. 1).

In our model, there are three types of entities: User, Gateway, and Satellite Network. Among them, Satellite Network does not directly take part in the authentication process, but only takes charge of forwarding authentication messages. Thus, only User and Gateway constitute the main part of end-to-end authentication.

User is able to request and start an end-to-end authentication. It could be either mobile or fixed, either large equipment or small hand-hold terminal. The process of authentication actually aims to construct a safe channel between two users for safe communication.

Gateway is a managing entity. It basically provides safe access management to the users registering to it; besides, it also accepts other users from other gateways and gets relevant information of these immigrant users to provide the same services.

3 End-To-End Authentication Protocol

End-to-end authentication protocol is the core part of a safe communication system. Mutual identity authentication and session-key negotiation could be conducted by the protocol, which would provide a confidential and integrated

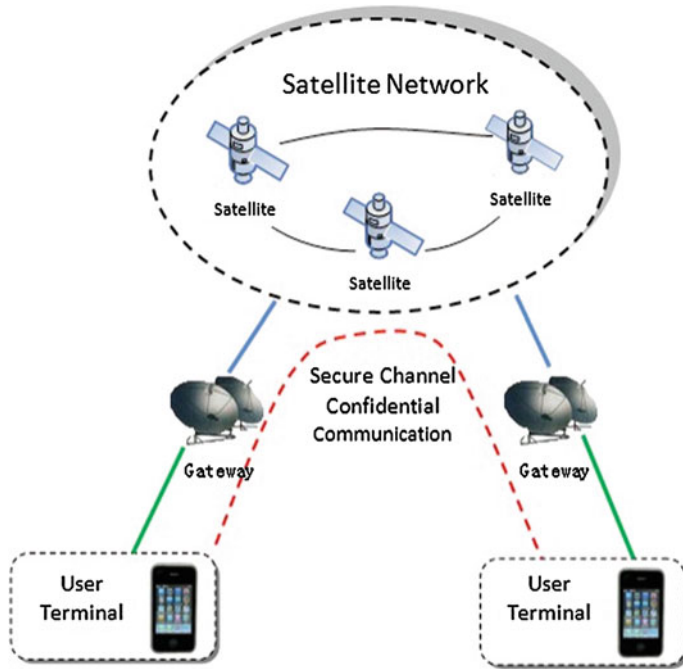


Fig. 1 Schematic diagram for end-to-end model

protection for the data transferring between two parties. This paper proposes an end-to-end authentication protocol based on the ECC public-key given an established Satellite Network model. The protocol is divided into three sub-protocols: initial authentication, authentication status inquires, and re-authentication.

3.1 Initial Authentication

Initial authentication is typical bidirectional identity authentication and private-key negotiation process. Given communication parties, User *A* and User *B* managed by different Gateways with User *A* by Gateway *A* and User *B* by Gateway *B*. Before launching a safe communication, an access authentication protocol has to be operated which is specified in Sect. 4. After a mutual authentication between User *A* and Gateway *A*, User *A* sends a session establishment request to User *B*. Gateway *A* forwards the request to Gateway *B*, and then, Gateway *B* sends the request down to User *B* to continue following negotiation. Fig. 2 depicts the process of initial authentication protocol.

User *A* is the call initiator and User *B* is the callee, and the protocol is described as follow:

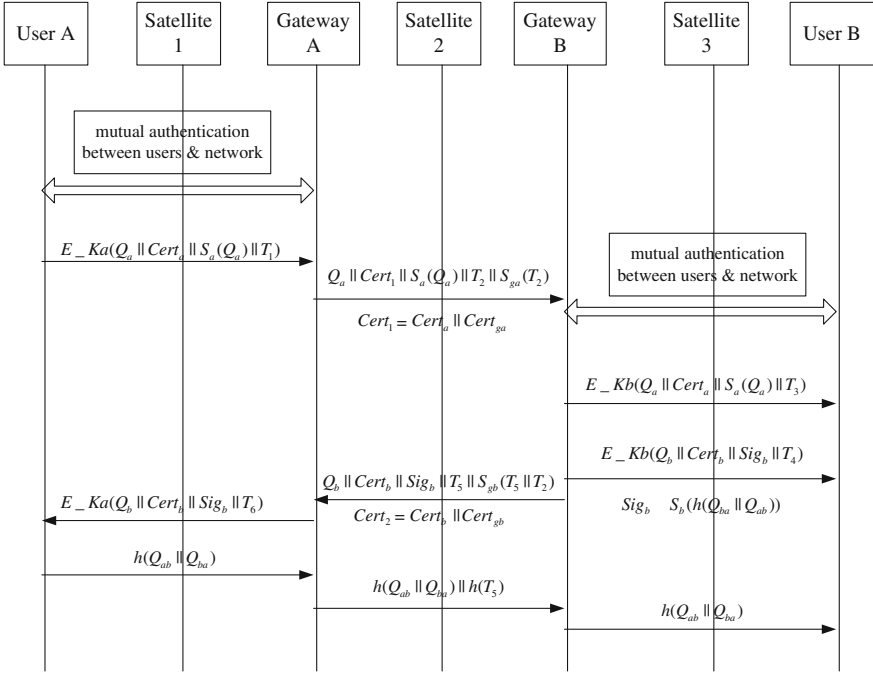


Fig. 2 Initial authentication

Step 1: Before User A initials a call request to User B, it has to execute an authentication of safe access to network to obtain the authority of access. User A shares private-key K_a together with Gateway A through private-key negotiation.

Step 2: User A sends call request $E_Ka(Q_a || Cert_a || S_a(Q_a) || T_1)$ to Gateway A.

After the authentication between User A and Gateway A, User A generates a random number R_a and time stamp T_1 , then computes the point $Q_a = R_a \times G$ on elliptic curve and achieves a signature $S_a(Q_a)$. Finally, User A sends call request $E(Q_a || Cert_a || S_a(Q_a) || T_1)$ to User B.

Step 3: Gateway A sends request to Gateway B with the message $Q_a || Cert_1 || S_a(Q_a) || T_2 || S_{ga}(T_2)$.

After receiving the request sent by User A, Gateway A deciphers $E(Q_a || Cert_a || S_a(Q_a) || T_1)$ and verifies the effectiveness of time stamp T_1 , $Cert_a$ and signature $S_a(Q_a)$. If the verification is passed, Gateway A modifies the request by generating time stamp T_2 together with a renewed signature $S_{ga}(T_2)$, and sends a message to Gateway B, $Q_a || Cert_1 || S_a(Q_a) || T_2 || S_{ga}(T_2)$, as User A's call request, among which $Cert_1 = Cert_a || Cert_{ga}$; else, it denies the request from User A.

Step 4: Gateway B sends request $E_Kb(Q_a || Cert_a || S_a(Q_a) || T_3)$ to User B.

Step 5: User B sends $E_Kb(Q_b || Cert_b || Sig_b || T_4)$ to Gateway B.

After receiving the message, User B deciphers it and verifies T_3 , $S_a(Q_a)$ and $Cert_a$. On passing the verification, User B continues processing the message by

storing $Cert_a$ and Q_a , responding to User A with $E_Kb(Q_b||Cert_b||Sig_b||T_4)$, among which T_4 is a time stamp, $Q_b = R_b \times G$, R_b is a random number, and with two parameters $Q_{ba} = R_b \times K_{pub}^a$, $Q_{ab} = K_{pri}^b \times Q_a$, and $Sig_b = S_b(h(Q_{ba}||Q_{ab}))$ implies the signature of User B's message abstract $h(Q_{ba}||Q_{ab})$.

Step 6: Gateway B sends $Q_b||Cert_b||Sig_b||T_5||S_{gb}(T_5||T_2)$ to Gateway A.

After receiving and deciphering the message, Gateway B verifies the effectiveness of T_4 and $Cert_b$. After the verification, it generates a time stamp T_5 together with a signature $S_{gb}(T_5||T_2)$. Finally, it responses to the call request message receiving from Gateway A with message $Q_b||Cert_b||Sig_b||T_5||S_{gb}(T_5||T_2)$.

Step 7: Gateway A sends to User A the message $E_Ka(Q_b||Cert_b||Sig_b||T_6)$.

After receiving the message, Gateway A verifies time stamp T_5 , signature $S_{gb}(T_5||T_2)$, and the public-key certificate of User B $Cert_b$. If the verification is passed, it will generate a response message for User A, $E_Ka(Q_b||Cert_b||Sig_b||T_6)$, among them T_6 is a time stamp and K_a is a private-key shared by User A and Gateway A.

Step 8: User A sends $h(Q_{ab}||Q_{ba})$ to Gateway A.

After receiving the response message, User A decipheres it and verifies time stamp T_6 , together with the signature $Sig_b = S_b(h(Q_{ba}||Q_{ab}))$ and $Cert_b$. If the verification is passed, it sends a connection assurance message $h(Q_{ab}||Q_{ba})$ to User B, among them $Q_{ba} = K_{pri}^a \times Q_b$, $Q_{ab} = R_a \times K_{pub}^b$.

Step 9: Gateway A sends $h(Q_{ab}||Q_{ba})||h(T_5)$ to Gateway B.

After receiving the connection assurance message, Gateway A sends it together with abstract value to Gateway B.

Step 10: Gateway B sends $h(Q_{ab}||Q_{ba})$ to User B.

After receiving the assurance message, Gateway B verifies $h(T_5)$. If the verification is passed, it sends $h(Q_{ab}||Q_{ba})$ down to User B. User B receives and verifies the correctness of $h(Q_{ab}||Q_{ba})$, and if the verification is passed, User A and User B finish the process of mutual authentication.

After mutual authentication, User A and User B both use the negotiated key to protect communication content, so the gateways only need to forward messages without considering deciphering it or other operations. By this method, an end-to-end authentication and communication scheme is achieved. The authentications and negotiations between those users under the management of the same gateway are similar, which will not be further illustrated.

3.2 Authentication Status Inquiry

Given User A and User B under management of different gateways (Gateway A and Gateway B). After conducting secrecy communication for a period of time, both the two ends also need to inquire whether to keep the status of authentication, and they will execute the supplementary protocol for authentication protocol—authentication status inquiry protocol. This protocol is started by one end (e.g.,

User A in Fig. 3). First of all, User A and Gateway A authenticate each other. Then, User A sends authentication status inquiry message to Gateway A to inquire whether there is a trustful status for User B in Gateway A. If there is, inquiry terminates with User A’s status inquiry counter increasing by one; or Gateway A sends a trustful status inquiry for User B to Gateway B. The following figure depicts the flow chart of an end-to-end authentication status inquiry protocol.

Here are the detailed descriptions of the protocol:

Step 1: User A starts the inquiry and Gateway is the callee.

Before sending an inquiry, User A has to execute the access authentication with service network to get access to it. After the authentication, User A shares the private-key K_a with Gateway A through negotiation and increases the status inquiry counter by one.

Step 2: User A sends inquiry request $E_Ka(Q_a || Cert_a || S_a(Q_a) || T_1)$ to Gateway A.

After their mutual authentication, User A generates random number R_a and a time stamp T_1 and calculates the point $Q_a = R_a \times G$ on elliptic curve with signature $S_a(Q_a)$. Finally, User A sends authentication status inquiry request on User B $E_Ka(Q_a || Cert_a || S_a(Q_a) || T_1)$ to Gateway A.

Step 3: After Gateway A receives the request sent by User A, the message $E_Ka(Q_a || Cert_a || S_a(Q_a) || T_1)$ is decrypted, and the time stamp T_1 , $Cert_a$ and signature message $S_a(Q_a)$ are verified. If the verification succeeds, we continue to deal with the message; otherwise, a request message with reject state is sent to

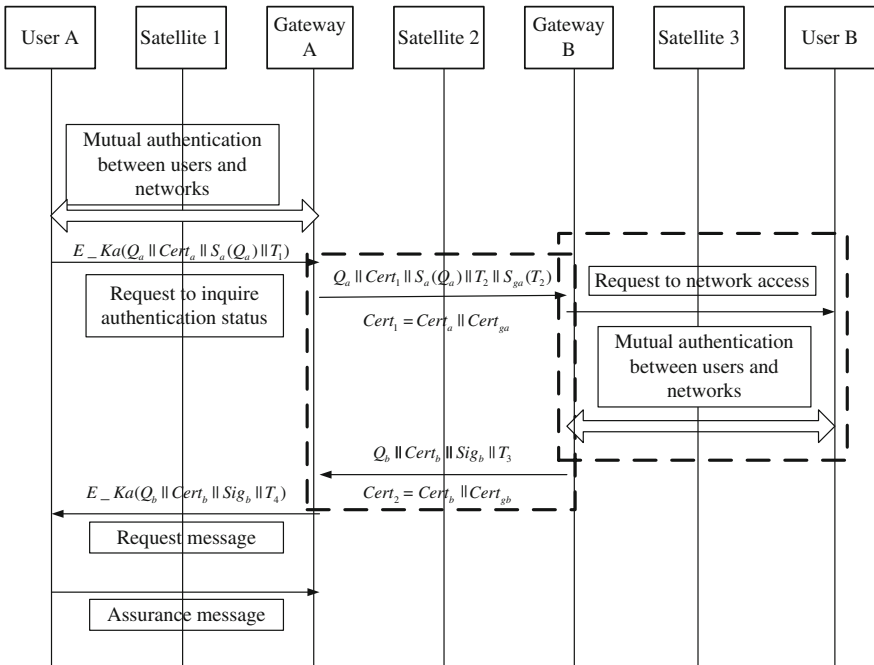


Fig. 3 Authentication status inquiry

User A. If Gateway A has the trust state of User B, an acknowledgment is sent to User A; otherwise go to Step 4.

Step 4: Gateway A is responsible for state inquiring to Gateway B.

Gateway A creates the time stamp T_2 , signs $S_{ga}(T_2)$, and sends $Q_a || Cert_1 || S_a(Q_a) || T_2 || S_{ga}(T_2)$ to Gateway B as the state inquiring message of User A, where $Cert_1 = Cert_a || Cert_{ga}$.

Step 5: After Gateway B receives the message, it verifies the authentication of certificate of User A $Cert_a$, $Cert_{ga}$ and signature $S_{ga}(T_2)$. If the verification succeeds, we continue to deal with the message; otherwise, a request message with reject state is sent to User A. If Gateway A has the trust state of User B, an acknowledgment is sent to User A; otherwise go to Step 6.

Step 6: Gateway B sends the network authentication request message to User B and asks for the network authentication of User B.

Step 7: User B generates the network authentication with Gateway B. If the verification succeeds, a response message of building connecting $Q_b || Cert_b || Sig_b || T_3$ is sent to Gateway A, where T_3 is the time stamp, $Q_b = R_b \times G$ (R_b is a random number), $Q_{ba} = R_b \times K_{pub}^a$, $Q_{ab} = K_{pri}^b \times Q_a$, $Sig_b = S_b(h(Q_{ba} || Q_{ab}))$ that is the signature of $h(Q_{ba} || Q_{ab})$. If the verification fails, a message denoting the failed-state inquiring is sent to Gateway A.

Step 8: Gateway A sends $E_Ka(Q_b || Cert_b || Sig_b || T_4)$ to User A.

Then, Gateway A receives the message, the time stamp T_3 , the authentication of certificate $Cert_b$ of User B and signature $S_{gb}(T_3 || T_2)$. If the verification succeeds, a state response message $E_Ka(Q_b || Cert_b || Sig_b || T_4)$ is created and sent to User A, where T_4 is the time stamp, and K_a is the common key of User A and Gateway A.

Step 9: User A sends the acknowledgment to Gateway A.

After User A receives the response message, it decrypts and verifies T_4 , signature $Sig_b = S_b(h(Q_{ba} || Q_{ab}))$, and the authentication certificate $Cert_b$ of User B. If the verification succeeds, an acknowledgment message is sent to Gateway A, where $Q_{ba} = K_{pri}^a \times Q_b$, $Q_{ab} = R_a \times K_{pub}^b$.

3.3 Duplicate Authentication

The trust relation exists between end-to-end users, which are reflected by the validity of shared key or deriving key. When the valid date will end or has been overdue, the end-to-end users need to have a new authentication and establish the trust.

The state when the key will expire or has been overdue is classified as follows:

The second operational state: when the key will expire, the encryption is not allowed and the decryption and verification are feasible;

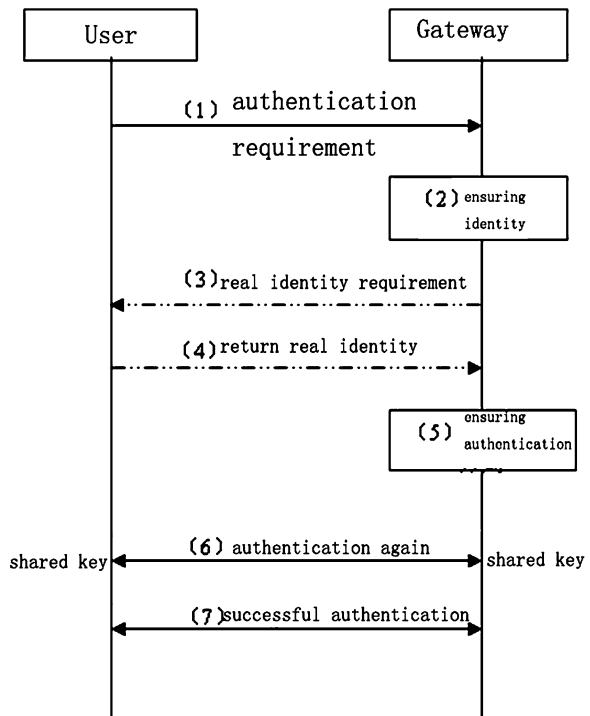
Withdraw state: The key has been overdue and the correspondence between the key and the real identification of the entity is released;

Destroy state: The record of the key is deleted.

Once the key is in the abnormal state during the end-to-end communication, the end-to-end double authentication is activated, which is described as follows (Fig. 4):

- Step 1: An end-to-end user, such as User A, activates the authentication procedure and carries the temporary identification and the way of last authentication in the request of authentication;
- Step 2: Gateway A labels the temporary identification according to the authentication way of the request message;
- Step 3: If the authentication way searched from the temporary identification of Gateway A disagrees with that of the user authentication, Gateway A asks for the real identification of the user;
- Step 4: The user returns its real identification;
- Step 5: Gateway researches the registration information from other gateways in terms of the real identification and ensures the both authentication ways;
- Step 6: User A and Gateway A implement the authentication by the selected way, and the shared key K_a is created.

Fig. 4 The end-to-end duplicate authentication protocol



4 The Protocol Security Analysis

The protocol references the literature [8] to verify by using semantic logic. Usually, the end-to-end session key is generated and allocated to the users by an authoritative third agency. Although this method is simple, it poses high requirement for the third agencies. The authentication and key-consulting protocol proposed by this paper is completed by the communication users and does not need the third agencies. So this protocol is of concise architecture, good safety, and high efficiency.

The digital certificate of the protocol realizes the authentication of the communication users [15], so the digital certificate is signed by the authentication center and cannot be fabricated. Moreover, the user signs the message by its key, which guarantees not only the resource of the message, but also the integrity of the information. After executing the protocol, both users can believe the identification each other.

5 The Simulation Analysis

This section arranges the safe service of end-to-end authentication including the simulation of voice and video services, and comparatively analyzes the end-to-end safe authentication service [9, 10] and the efficiency of the proposed distributed end-to-end safe authentication.

The Walker constellation of satellite network is composed of 20 MEO satellites; there are 4 planes, each plane with 5 satellites. The orbital inclination is 75 degree, orbital altitude is 14,163 km, Rann angle is 360°, and the phase off is 1. Three gateways are distributed in china and 11 users are uniformly allocated to the whole word. The simulation platform is Pentium IV 2.8 G CPU, 512 M memory, and OS is Windows XP.

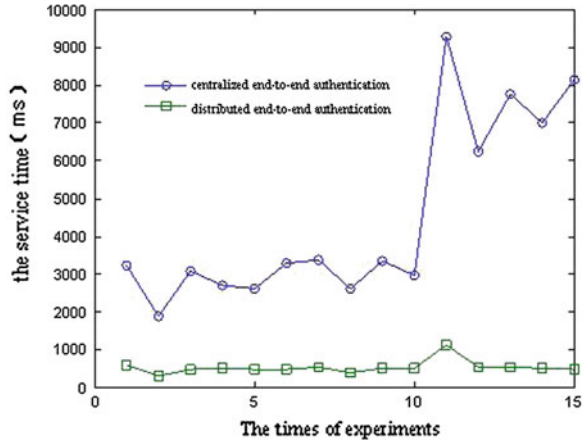
The simulation mainly focuses on the process of end-to-end safe service authentication. After the experiments, we analysis the time generated, respectively, by the centralized safe authentication and distributed safe authentication according the experimental results.

5.1 Voice Service

The voice service experiment simulates the service of 11 users. For every service, the arbitrary two users create a service process (total 110 times). The Inter-plane EBR is $EBR = 10^{-4}$, the up-link EBR is $EBR = 10^{-5}$, and down-link EBR is $EBR = 10^{-5}$. The average service duration is 10 s.

Figure 5 shows the experimental results of voice simulation. It can be seen that the time cost by distributed end-to-end authentication is remarkably less than the

Fig. 5 The experiment results of safe voice simulation



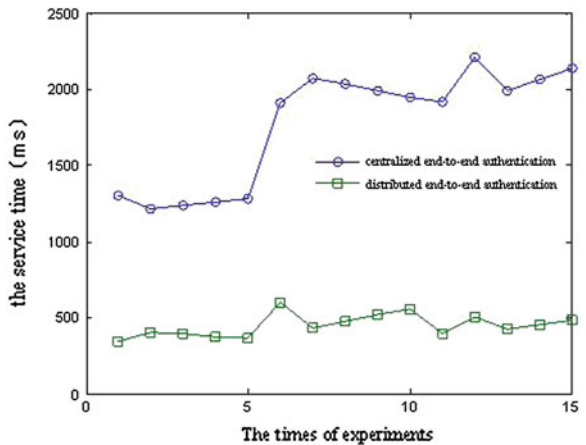
centralized scheme. The reason may lie in that the distributed scheme allocates many authentication works to some different gateways, which improves the authentication efficiency and reduces the authentication time.

5.2 Video Service

The experiment of video service simulates the service of 11 users. For every service, the arbitrary two users create a service process (total 110 times). The Inter-plane EBR is $EBR = 10^{-5}$, the up-link EBR is $EBR = 10^{-6}$ and down-link EBR is $EBR = 10^{-6}$. The average service duration is 10 s.

Figure 6 shows the experimental results of video simulation. The average time cost by distributed end-to-end authentication is about 500 ms. However, the time

Fig. 6 The experiment results of safe video simulation



cost by the centralized scheme is more than 1,000 ms. It is clear that the distributed end-to-end authentication has better efficiency. If we comparatively analyzes the voice and video simulation results, it can seen that the time cost by two services is basically similar, which is because the end-to-end authentication is not relate the service. Moreover, the data size of the authentication packet is small and the bit error rate of the channel has a less effects.

6 Conclusion

This paper analyzes the limitation of conventional end-to-end authentication technology of satellite networks, proposes the distributed end-to-end authentication technology, which uses the management and trust relation of the gateways to effectively reduce the space transmission effects of end-to-end authentication process, and improves the efficiency of the end-to-end authentication. Besides, this paper discusses the safety of the proposed end-to-end authentication protocol through the simulation.

Acknowledgments The work is jointly supported by the National Natural Science Foundation of China (Grants No.: 60973145, 61004021).

References

1. Miller SP, Neuman BC, Schiller JL, Saltzer JH (1987) Section E.2.1: Kerberos authentication and authorization system. MIT Project Athena, Cambridge, Massachusetts (21 Dec 1987)
2. Kohl J, Neuman S (1993) The Kerberos network authentication service. RFC 1510
3. Chang YF, Chang CC, Shiu CY (2005) An efficient authentication protocol for mobile satellite communication systems. *ACM SIGOPS Oper Syst Rev* 39(1):70–84
4. Chadwick DW, Otenko O (2002) The PERMIS X. 509 role based privilege management infrastructure. In: *Proceeding of SACMAT' 02*, pp 1–9
5. Diffie W, Hellman ME (1976) New directions in cryptography. *IEEE Trans Inf Theor* 14(5):111–222
6. Xu ZB, Ma HT (2007) Design and simulation of security authentication protocol for satellite network. *Comput Eng Appl* 43(17):130–132
7. Xu Y (2009) Design and simulation of security authentication protocol for satellite networks. Master degree thesis of Chinese Academy of Sciences
8. Xu ZB, Ma HT (2007) Design and simulation of security authentication protocol for satellite network. *Comput Eng Appl* 43(17):130–132
9. Sandirigama M, Shimizu A, Noda MT (2000) Simple and secure password authentication protocol. *IEICE Trans Commun* E83–B(6):1363–1365
10. Yuan D, Fan PZ (2002) A secure dynamic password authentication scheme. *J Sichuan Univ (Nat Sci)* 39(2):228–232
11. Halevi S (2007) Invertible universal hashing and the TET encryption mode: CRYPTO 2007. Springer, Berlin, pp 412–429
12. Bellare M, Canatti R, Krawczyk H (1996) Keying hash function for message authentication—CRYPTO'96. Springer, Berlin, pp 1–19

13. Carter J, Wegman M (1981) Universal classes of hash functions. *J Comput Syst Sci* 22(2):143–154
14. Stinson RD (1994) Universal hashing and authentication codes. *Des Codes Crypt* 4(4):369–380
15. Yong Xu (2010) Research on key technologies of end-to-end secure transmission systems. Master degree thesis of Beijing Jiaotong University

Current Status, Challenges and Outlook of E-health Record Systems in China

Xiangzhu Gao, Jun Xu, Golam Sorwar and Peter Croll

Abstract China (the mainland) is a developing country with a large population over a vast area, where healthcare services are not balanced. The Chinese government faces a mammoth task to provide health and medical services to the population. About eight years ago, the government defined its direction to electronic health (e-health). Since then, hundreds of e-health record systems have been developed, but the systems are still in their infancy. For the establishment of matured e-health record systems, this paper examines current status of China's development of e-health record systems, identifies the challenges encountered by the development and analyses the outlook of future systems in China.

Keywords Health care · E-health record systems · EHR · EMR · China

1 Introduction

From 2000 to 2010, Chinese population increased by 5.84 %, reaching 1.34 billion according to the sixth national census [1]. To meet the needs of the health care by such a large population and to improve healthcare services, China increased its annual health expenditure dramatically by 335 % from CN¥460 billion in 2000 to CN¥2000 billion in 2010, according to Ministry of Health (MOH) [2]. Over these years, the healthcare infrastructure in China has been well developed. In 2011, there were 954,389 health institutions, including 21,979 hospitals, 918,000 grass roots/community healthcare centres and 11,926 specialised public health institutions.

X. Gao (✉) · J. Xu · G. Sorwar · P. Croll
Southern Cross University, Lismore, NSW, Australia
e-mail: xiangzhu.gao@scu.edu.au

However, as China is a developing country with a large population, China's health expenditure per capita is much lower than that of developed countries. In 2008, China's health expenditure per capita was US\$146, contrasting United States's US\$7,164, France's US\$4,966, Germany's US\$4,720, Canada's US\$4,445, Australia's US\$4,180, United Kingdom's US\$3,771 and Japan's US\$3,190 [2]. To further improve its health care, China must increase efficiency and decrease costs of its healthcare services, besides the increase in government budget.

The Chinese government laid a focus on healthcare sector in the 11th Five-Year Plan (2006–2011). Included in the Plan is a 3521 project, where the '2' indicates two systems: electronic health record (EHR) system and electronic medical record (EMR) system [3–5]. A health record is the information of health conditions about a healthcare recipient [6], while a medical record is the information of hospital clinic and therapy about a medical care recipient [7]. EHR is collected, stored and shared by regional community health service centres, while EMR is recorded or generated, stored and shared by hospitals. It was planned that the EHR system would cover 70 % of urban residents [5].

The government has invested significantly into e-health initiatives. In its new medical reform plan, CN¥850 billion was allocated for the period of 2009–2011 to modernise healthcare services with the introduction of EMR systems and digital hospitalisation, aiming at providing basic medical service to all Chinese citizens. The plan also allocated CN¥3.9 billion for EHR initiatives in 2010. For the operation of EHR and EMR systems, CN¥53.4 billion was allocated for the period of 2010–2012 to invest in medical IT and e-health infrastructure. It was estimated that the total investment in 2012 would be CN¥40 billion [4, 5, 8].

The Chinese government leads the development of EHR and EMR systems. The MOH approved the establishment of professional committees for research on and development of standards and specifications. The Chinese Hospital Association Management Committee (CHIMA) focuses on the preparation of management specification and regulations for digital hospitalisation standards [9]. The HISPC of the MOH prepared a series of guidebooks, standards and specifications for both EHR systems and EMR systems. In addition, the government has developed supportive e-health policies, including national e-health policy, national multiculturalism policy for e-health, national telemedicine policy and e-government policy.

China's e-health system development is significantly affected by its economy, administration, geography, demography and culture. Economic development in China is extremely unbalanced. Urban development is faster than rural development, and growth in eastern regions is faster than that in western regions. The Chinese population is rapidly ageing, due to a lower mortality rate and the one-child policy, which was introduced in 1979. In 2010, rural residents were over 674 million (50.32 %); mobile population was over 261 million (19.07 %); and elderly residents above 60 years old were over 177 million (13.26 %) [1]. These features have led to healthcare challenges for the Chinese government.

In China, community health service centres play an important role in prevention, medical care, rehabilitation and health promotion. The centres are non-profit institutions, and their service targets are women, children, elderly people,

chronically sick and disabled people centres and poor residents in the community. Communities are partitioned based on the administrative areas where households are registered according to China's household registration system. A household registration record includes the members of a family as residents of an administrative area and identifies each member's information of name, gender, date of birth and the relationship with the householder (parent, spouse, etc.). The household registration records, and the roles and targets of the community health service centres affect the development of EHR systems.

2 Current Status

According to CHIMA [9], computerisation in hospitals in China started in late 1980s. Some economically advanced cities started to develop regional e-health record systems in 2002 [10]. In this paper, however, we shall only cover relevant efforts and results since 2005. The development of EHR/EMR systems in China is marked with the following milestones.

- In 2005, the fifth plenum of the 16th Communist Party Central Committee passed its 'suggestion' for the 11th Five-Year Plan [11]. The Plan directed the development of EHR system and EMR system.
- In 2006, the Health Information Standardisation Professional Committee (HIS-PC) was established by the MOH for standards creation, administrative certification and application promotion [12].
- In 2009, the MOH issued the Guidebook for Creating Regional Health Information Platform Based on EHRs. The Guidebook describes operation flow, and information and functionality of the platform, providing IT vendors with clear system requirements [13].
- In 2009, a series of standards and specifications were published by MOH, including Basic Structure and Data Standards for Health Records [14], Basic Dataset Compilation Specification for Health Records [15], Common Data Elements in Health Records [16] and Basic Dataset of Personal Information [17].
- In 2009, Beijing, Shanghai, Guangzhou, Xiamen and Hangzhou, which are economically advanced cities in China, saw their initial achievements in EHR system initiatives [10].
- In 2010, Shanghai, Zhejiang, Anhui, Chongqing and Xiangjiang, which are located from east through west in China, participated in testing EHR/EMR systems [5], financially supported by MOH.
- In 2011, the Guidebook for Creating Integrated Administration Platform for Health Records was released by MOH. This Guidebook includes functional requirements and design concepts for the platform, which provides health administrators and decision makers with information support and service [18].

- In 2011, the MOH published the Technical Solution for Developing Hospital Information Platform Based on EMRs [19].
- In 2012, the MOH published the Specification for Sharing Documents of Health Information [6] and the Specification for Sharing Documents of EMRs [7].

In planning for the EHR/EMR systems, China studied international experiences from United States, United Kingdom, Canada, Australia, etc. For example, China's EMR systems adopt the HL7 Standard [19], which takes an important position in US standard system and has become an international standard for medical information exchange [20]. Because of the complexity of EHR systems, Australia has taken three stages to reach a national system: MediConnect, which was a pilot system [21]; Health Connect, which consisted of seven regional systems [22]; and personal controlled electronic health record (PCEHR) system, which is a national system [23]. China follows the same stages.

By 2009, China had completed its pilot test with the EHR systems of major cities in China [10]. In this initial stage, some cities or provinces developed their own standards. Based on these systems, the MOH developed standards and specifications for the following development stages. Until 2011, 120 EHR systems had been completed, 40 systems were under development, and 100 systems were in planning phase [18]. The 2009 Guidebook [13] required that by 2011, 50 % urban residents and 30 % rural residents will have their e-health records, and by 2020, all residents will have their own e-health records. The Minister of MOH declared on 27 February 2012 that 900 million residents had created their health records, accounting for 66 % of the national population, and more than 50 % of residents had created their e-health records (chinanews.com 2012) [24].

For promotion of the EHR systems, the Handbook for Using Standardised Resident EHR System was developed. According to the handbook, an existing standardised EHR system consists of five modules: children's health, women's health, disease management, disease control and medical service, which are basically the targets of community health service centres.

In February 2012, the Ministry of Science and Technology of China accepted the Research on National Digital Health Key Techniques and Regional Demonstration and Application [24]. The two specifications in 2012 [6, 7] lay a foundation for the national EHR/EMR systems, and the successful research indicates the immediate step into the third development stage.

Regarding EMR systems, there exist many stand-alone and scattered efforts across the country. IT vendors have worked with hospitals to provide solutions for EMR systems. Successful examples are as follows:

- IMB has developed a series of solutions, including Clinical and Medical Record Analysis and Sharing System (CHAS) for a hospital in Guangdong [25] and Cloud Solution for sharing clinical data [26].
- Cisco [27] developed an Integrated Solution for Digital Hospitals for Peking University People's Hospital.
- Dell provided its solution of virtual technology to help Xiamen University Zhongshan Hospital upgrade its existing systems [28].

Hospitals develop EMR systems without a consideration about sharing information with EHR systems. Xiamen is the only city that has successfully attempted the association of an EHR system with an EMR system in the first stage. In Xiamen's system, health information, medical treatment records and check-up results of a resident since birth are recorded in a data repository. This system covers 95 medical institutions and 60 % of the regional residents [29].

Health Cards is another major project relating to connecting EHR and EMR. The card system will link hospitals, public health institutions and insurance operations across China. Currently, the system is piloted in Henan, Liaoning, Guangdong and Inner Mongolia [4].

Regional healthcare information network (RHIN) was an attempt to share data between communities. RHIN utilised data centres and telecommunication networks to share data and clinical services among geographically dispersed communities [30].

3 Challenges and Issues

3.1 Insufficient Fund

E-health projects are expensive. United States invested AU\$28 billion (AU\$1320 per capita) in the Barack Obama Plan in 5 years, and Australia has already spent AU\$466.7 million (AU\$21 per capita) on the first release of the PCEHR system in the last 2 years [31]. It is estimated that the required investment in hardware/software for normal operation in China is CN¥26 billion (CN¥20 per capita) [10, 32]. If the hardware/software accounts for 20 % of the development cost [33], the development cost would be CN¥130 billion (CN¥100 or AU\$17 per capita). Generally speaking, the annual cost to maintain an e-commerce or e-government system in its operation phase is between 50 and 200 % of the cost for the development phase [33]. According to this rule, China would spend CN¥65–CN¥260 billion annually to maintain the systems. Although the total healthcare expenditure was CN¥2000 billion in 2010, the government contribution is CN¥573 million [1]. Obviously, government healthcare budget cannot provide sufficient fund for system operation.

3.2 Lack of Unified Planning and Governance

Before 2010, when the five cities were selected and supported by MOH to test EHR/EMR systems, all system developments had happened at regional level without centred governance. Even though there are some MOH documents for EHR/EMR systems, they are not clear and difficult to follow in practice [34].

In 2004, the Chinese Health Information Society established the Health Information Standardisation Professional Committee (CHIS-HISPC). In spite of the existence of CHIS-HISPC, MOH established its HISPC in 2006. The two committees have partly overlapped roles and/or functionality.

Currently, hundreds of systems are developed or operating at regional levels across the country with no or little data sharing, which is an essential requirement for EHR/EMR systems. Some regional governments have developed local standards, for example the Technical Specification of Access Interface for Sharing Health Data by Zhejiang Bureau of Quality and Technical Supervision [35]. Same work is repeated at low quality levels. These systems are not effective for controlling pandemic diseases such as SARS, which spread in China in 2003.

Because of the lack of unified national approach, clear planning and governance framework, including assessment and evaluation, the quality of the EHR/EMR systems and the productivity of the development are low. A large amount of resources are wasted.

3.3 Issue of Two Systems

In China, EHR system and EMR system are regarded as different systems and developed in parallel individually. However, much information included in EHR system can be obtained from EMR system [19]. Currently, standards or specifications are developed for each of the systems. The potential incompatibility will hinder data transmission between the systems or breach data integrity. This is a serious problem for health or medical data.

3.4 Lack of Legislation for the Use of EMR/EHR

EHR/EMR systems should ensure data secrecy and availability and protect data integrity within a system, when data are transmitted between systems in a country or between countries [36]. Private data should be secret, useful data should be available, and critical data must be correct and accurate. Even though security techniques such as authentication and encryption can address these issues, legislation, including privacy policies, is necessary to regulate system operation (e.g. authentication and security key management) and the use of the system (e.g. 'meaningful use'). For the development and implementation of EMR/EHR systems in China, there is a lack of policy and regulation for using, recording and storing personal health information [10]. However, existing policies or regulations of other countries cannot be copied for China's EHR/EMR systems because of the difference in culture, law, administration, etc. For example, the household registration record, which is personally identifiable data, is not regarded as private information in China.

3.5 Issue of Compatibility with International Standards

Even though China has issued its standards, major improvements have to be made by comparing with international standards. For example, the issued trial version of Basic Structure and Data Standards for Health Records [14] has some critical deficiencies in terms of the information on privacy and security, support for different data types and reference data, mechanism to support easy additions and extensions to various medical domains and organisations and relational attributes for data elements [37].

3.6 Issue of E-Health Talent, Skills and Research

We noticed that major health IT solution providers in Chinese market are IBM, Cisco, Dell and Microsoft. Although there are some Chinese firms providing solutions, they fail in providing their own successful cases. Health/medical informatics education and training in China cannot meet China's need [38]. There is a lack of skilled health IT professionals to develop health IT solutions [39] and the talent to combine IT knowledge/skills, health care, nursing and public health [34]. According to World Health Organisation [36], China has been working on developing e-health capacity building in recent years via ICT training and education for health students and health professionals. It is expected that the problem of lack of talent and skill will be solved in a few years. We also noticed that it is difficult to obtain literature of theory on e-health in China. This indicates a lack of theory support for developing e-health information systems in China [32, 40].

3.7 Need for EHR System Education

According to a research in 2011 [10], data for EHR systems are mainly collected in urban communities and rural villages from women, children, elderly people, chronically sick and disabled people, who are the main targeted users of China's EHR systems. Compared with others, these groups of people are averagely at lower levels in IT literacy. There is a need for EHR system education among these people.

4 Outlook and Conclusion

4.1 A Single EHR/EMR National System

China has experienced the stage of pilot systems and the stage of regional systems. Although there are hundreds of regional EHR/EMR systems, these systems will have to be abandoned or upgraded soon because of the issues discussed in the previous section, such as poor or little capability to share data with other systems and high maintenance costs. However, MOH has developed a series of standards and specifications for national EHR and EMR systems in this year and will continue investing in the development. China will step into the third stage of national systems immediately. Australia has experienced the same 3 stages and has released its first version of national PCEHR system after 2 years' effort. It is expected that China will release its initial version of a national EHR system in 2015, according to Australia's experience. As the standards and specifications for EHR and EMR systems are developed by the same organisation, which will continue to lead the development of the systems, there will be one single EHR/EMR national system in China eventually.

4.2 Adoption of EMR Systems in Hospitals

In 2008, 80 % of hospitals had implemented a hospital information system of a kind. According to MOH [2], there are 21,979 hospitals in China. This means that 17,583 hospitals have installed or used an information system for medical purposes or supporting activities. Because of the diversity of the systems, although some of them may be the same, maintenance costs are high. We noticed that IMB has provided Cloud Solution and Dell has provided virtual technology for EMR systems. The solution and technique provide flexible computing and enable effective management of EMR. Because of the investment in EMR systems and the Technical Solution [19] and Specification 2012 [7], it is expected that by 2015, all level 3 (the highest ranking in 3 levels) hospitals would have established a unified EMR database [4].

4.3 Mobile Health

Currently, China has not taken an initiative to access health records using mobile devices (mobile health or m-health). However, the mobile population and mobile phone population determine the significance of m-health in China. Because of the unbalanced economy, many people from impoverished regions go to more urban and prosperous coastal regions in search for work. These people construct the majority of

the mobile population. They rely on mobile devices more than on normal computers to access the Internet. In 2011, mobile population was over 261 million [1]. In May 2012, China has 400 million mobile Web users as world's top smartphone market [41]. There is a need for EHR systems that allow mobile access.

Acknowledgments This research is supported by Australian Government under the Australia–China Science and Research Fund.

References

1. National Bureau of Statistics of China (2011) 2010年第六次全国人口普查主要数据公报, 28 Apr 2011. Online available http://www.stats.gov.cn/tjfx/jdfx/t20110428_402722253.htm. Accessed 02 Oct 2012
2. MOH (2012a) 2012 China health statistics (executive summary) (in Chinese), Centre for Statistics Information, Ministry of Health, China, 6 June 2012. Online available at <http://www.moh.gov.cn/publicfiles/business/htmlfiles/mohwsbwstjxxzx/s9092/201206/55044.htm>. Accessed 26 Sep 2012
3. KPMG China (2011) China's 12th five-year plan: healthcare sector, KPMG China, May 2011, Online available <http://www.kpmg.com/cn/en/IssuesAndInsights/ArticlesPublications/Documents/China-12th-Five-Year-Plan-Healthcare-201105-3.pdf>. Accessed 12 Mar 2012
4. New Zealand Trade and Enterprise (2012) Expert guide: health IT in China, Market Profile July 2012, Online available at: <http://www.nzte.govt.nz/explore-export-markets/market-research-by-industry/Information-and-communication-technologies/Documents/North-Asia-Health-IT-in-China-July-2012.pdf>. Accessed 2 Sep 2012
5. Teng G, Li M, Li V (2012) The digital revolution in China. PACS and Networks, 8th Annual MIIT Conference, May 11, 2012, Toronto, Canada, Online available <http://miircam.com/miit2012/2-Li.pdf>. Accessed 25 Sep 2012
6. MOH (2012c) 健康档案共享文档规范. Online available at http://www.ahwst.gov.cn:5300/xxgkweb/showGKcontent.aspx?xxnr_id=10290. Accessed 12 Aug 2012
7. MOH (2012b) 电子病历共享文档规范. Online available at http://www.ahwst.gov.cn:5300/xxgkweb/showGKcontent.aspx?xxnr_id=10291. Accessed 12 Sep 2012
8. Research in China (2011) China medical information system industry report 2011, Market Publishers, October 2011, Online available <http://ebookbrowse.com/china-medical-information-system-industry-report-2011-pdf-d336971697>. Accessed 20 Sep 201
9. CHIMA (2008) The white paper on China's hospital information systems. Online available at <http://www.chima.org.cn/pe/DataCenter/ShowArticle.asp?ArticleID=612>. Accessed 01 Oct 2012
10. Huang W (2011) Generating development strategies of electronic health records in china based on stakeholders analysis (in Chinese), Master Thesis. Research Institute of Medical Informatics, China
11. Naughton B (2005) The new common economic program: China's eleventh five year plan and what it means, China Leadership Monitor, No. 16
12. Wang Y (2012) 我国卫生信息标准化现状与发展研究, 14 June 2012. On-line available at http://md.tech-ex.com/html/2012/article_0414/18664.html. Accessed 02 Oct 2012
13. MOH (2009b) 基于健康档案的区域卫生信息平台建设指南. Online available at <http://www.chima.org.cn/pe/Article/ShowArticle.asp?ArticleID=764>. Accessed 09 Aug 2012
14. MOH (2009d) 健康档案基本架构与数据标准. Online available at <http://www.chima.org.cn/pe/Article/ShowArticle.asp?ArticleID=763>. Accessed 09 Aug 2012
15. MOH (2009e) 健康档案基本数据集编制规范. Online available at <http://www.chima.org.cn/pe/Article/ShowArticle.asp?ArticleID=762>. Accessed 09 Aug 2012

16. MOH (2009c) 健康档案公用数据元标准 Online available at <http://www.chima.org.cn/pe/Article/ShowArticle.asp?ArticleID=761> .Accessed 09 Aug 2012
17. MOH (2009a) 个人信息基本数据集标准. Online available at <http://www.chima.org.cn/pe/Article/ShowArticle.asp?ArticleID=760>. Accessed 09/08/2012
18. MOH (2011b) 卫生综合管理信息平台建设指南. Online available at <http://www.moh.gov.cn/publicfiles/business/cmsresources/wsb/cmsrsdocument/doc11824.pdf>. Accessed 09 Aug 2012
19. MOH (2011a) 基于电子病历的医院信息平台建设技术解决方案. Online available at <http://www.moh.gov.cn/publicfiles/business/htmlfiles/mohbgt/s6694/201103/51091.htm>. Accessed 08 Aug 2012
20. Zhang, Y (2012b) 国外卫生信息标准化现状及发展趋势, 14/06/2012, Online available at http://md.tech-ex.com/html/2012/article_0614/18286.html. Accessed 02 Oct 2012)
21. Medicare (2012) MediConnect, 24 Aug 2012. Online available at <http://www.medicareaustralia.gov.au/provider/patients/medicconnect.jsp>. Accessed 27 Sep 2012
22. Department of Health and Aging (2009) Health Connect Evaluation. Online available at [http://www.health.gov.au/internet/main/publishing.nsf/Content/B466CED6B6B1D799CA2577F30017668A/\\$File/HealthConnect.pdf](http://www.health.gov.au/internet/main/publishing.nsf/Content/B466CED6B6B1D799CA2577F30017668A/$File/HealthConnect.pdf). Accessed 27 Sep 2012
23. NEHTA (2011) Concept of operations: relating to the introduction of a personally controlled electronic health record system. Online available at [http://www.yourhealth.gov.au/internet/yourhealth/publishing.nsf/Content/PCEHRS-Intro-toc/\\$File/Concept%20of%20Operations%20-%20Final.pdf](http://www.yourhealth.gov.au/internet/yourhealth/publishing.nsf/Content/PCEHRS-Intro-toc/$File/Concept%20of%20Operations%20-%20Final.pdf). Accessed 26 Sep 2012
24. Chinanews.com (2012) 中国卫生数字化提速 居民电子健康档案建档率过半, 27 Feb 2012 Online available at <http://www.chinanews.com/jk/2012/02-27/3701936.shtml>. Accessed 03 Oct 2012
25. Zhi Bang Software Technology (2011) IBM 合作创新先进的医疗信息共享与分析技术. Online available at <http://www.zbintel.com/wz/52317111.htm>. Accessed 03 Feb 2012
26. IBM 2012, IBM—医疗云解决方案, 16 July 2012. Online available at <http://www.cloudguide.com.cn/news/show/id/1755.html> Accessed 03 Oct 2012
27. Cisco (2012) 北京大学人民医院携手思科通过数字化医院建设全面提升医护体验, 12 Mar 2012. Online available at http://www.cisco.com/web/CN/aboutcisco/news_info/china_news/2012/03_12.html. Accessed 03 Oct 2012
28. e800.com.cn (2011) 戴尔助力厦门大学中山医院全面升级IT系统, 30 Dec 2011. Online available at <http://www.e800.com.cn/articles/2011/1230/501128.shtml>. Accessed 03 Oct 2012
29. Whatsonxiamen.com (2010) 'Xiamen takes the lead in china in electronic health record system', 25 Aug 2010. Online available at: <http://www.whatsonxiamen.com/news14185.html> . Accessed 25 Sep 2012
30. Zita K (2009) China healthcare ICT: reinventing China's national healthcare systems through electronic medical records, telecom networks and advanced IT services. *J Emerg Knowl Emerg Market* 1(1):47–54
31. NEHTA (2012) E-Health and the implementation of the PCEHR, Online available at. Accessed 29/09/2012
32. Shuai P, Tang D (2010) Electronic health record executive and management (in Chinese). *Chin Sci Technol Res Rev* 42(5):55–60
33. Schneider G (2011) *Electronic commerce*, 9th edn. Course Technology, Thomson Learning, Boston
34. Zhang X (2012) International and domestic research on electronic health record. *Int J Med Inf* 34(3):236–239
35. Zhejiang Bureau of Quality and Technical Supervision (2011) Technical specification of access interface for sharing health data (in Chinese). Online available at <http://www.zjbs.gov.cn/html/main/bztzggView/241958.html>. Accessed 08 Aug 2012

36. World Health Organization (2011) ATLAS: eHealth Country Profiles 2010, World Health Organization 2011, Geneva. Online available http://www.who.int/goe/publications/ehealth_series_vol1/en/index.html. Accessed 27 Mar 2012
37. Xu W, Guan Z, Cao G, Zhang H, Lu M, Li T (2011) Analysis and evaluation of the electronic health record standard in China: a comparison with the American national standard ASTM E 1384. *Int J Med Inf* 80:555–561
38. Zhang Y, Xu Y, Shang L, Rao K (2007) An investigation into health informatics and related standards in China. *Int J Med Inf* 76:614–620
39. Wikipedia (2012) Health informatics in China, Wikipedia.org, 14 Aug 2012. Online available at http://en.wikipedia.org/wiki/Health_informatics_in_China. Accessed 25 Sep 2012
40. Guo H, Dai T, Hu H, Huang W (2010) Current status, hot topics, and trends of electronic health record: analysis based on PubMed data (in Chinese). Online available <http://ziyuan.iyyi.com/source/show/1404176.html>. Accessed 02 Sep 2012
41. MobiThinking (2012) China passes 1 billion mobile subscribers, passes 400 million mobile Web users and overtakes US as world's top smartphone market. Online available at <http://mobithinking.com/blog/china-top-mobile-market>. Accessed 04 Oct 2012

Singularity Analysis of the Redundant Robot with the Structure of Three Consecutive Parallel Axes

Gang Chen, Long Zhang, Qingxuan Jia and Hanxu Sun

Abstract A method based on matrix partitioning is presented to analyze the singularity of a redundant robot with the structure of three consecutive parallel axes. Unchanging robot singularity, an appropriate reference system and a reference point of the end-effector are chosen to simplify the analytical form of the Jacobian matrix. Thereafter, the reason why a 3×3 zero matrix exists in the simplified Jacobian matrix is analyzed specifically, and according to this characteristic, the Jacobian matrix is reconstructed by matrix transformation. On this basis, the Jacobian matrix is partitioned into four submatrixes whose degradation conditions will be discussed. In the light of these degradation conditions, the singularity conditions and singular directions of the redundant robot can be obtained. The correctness of the proposed singularity analysis method for the redundant robot with the structure of three consecutive parallel axes is verified through calculation examples.

Keywords Structure of three consecutive parallel axes · Redundant robot · Singularity analysis · Matrix partitioning

1 Introduction

Singularity is the inherent kinematic characteristic of robots. At a singularity configuration, the end-effector of robot loses the ability to move along a certain direction, which is called the singular direction. In this case, if there is still a velocity component of the end-effector in the singular direction, the joint angular velocities will become unacceptably large, and the end-effector will deviate from

G. Chen (✉) · L. Zhang · Q. Jia · H. Sun
School of Automation, Beijing University of Posts and Telecommunications,
Beijing 100876, China
e-mail: buptcg@gmail.com

the expected trajectory. What's worse, it is possible that a robot will be out of control. Therefore, in order to ensure the stability and reliability of robot system, it is necessary to do the research on singularity analysis and path planning for singularity avoidance. Robot singularity analysis, as the basis of singularity avoidance, can determine the singularity conditions. For non-redundant robots, the determinant value of Jacobian matrix can directly determine robot singularity conditions. However, for redundant robots, because its Jacobian matrix is not square, the singularity analysis becomes much more complex.

Much effort in research community has been paid on dealing with the singularity analysis of redundant robots. Whitney [1] determined singularity conditions by calculating the determinant value of matrix JJ^T . Singular configurations occur when the determinant value is equal to zero. Although the matrix formed by JJ^T is square, the determinant value expression is quite complicated, making it very difficult to obtain the analytical expression of singularity conditions. Podhorodeski et al. [2] proposed using six-joint subgroups of Jacobian matrix to determine the singularity conditions of redundant robots. Conditions that cause the determinant values of all possible six-joint subgroups to simultaneously equal zero are singularity conditions. The method works well, but the entire solving process is much complex. Nokleby et al. [3, 4] used the properties of reciprocal screws to determine the single-DOF-loss and multi-DOF-loss conditions of redundant robots and also work out the singular directions. This method is applied to the singularity analysis of Canadian Space Agency (CSA) Canadarm2 [5]. The singularity conditions of redundant robots can be easily obtained and its computational efficiency is high, while the principle is too abstract to understand clearly. For a 7-DOF robot with the structure of the last three axes perpendicular to each other, Waldron et al. [6], based on matrix partitioning, extended the segmented 3×4 submatrix to a 4×4 submatrix with an additional relationship. And then, the singularity conditions can be achieved through the determinant value of the square submatrices. Due to the introduction of an additional relationship, the method may encounter algorithm singular problems. Cheng et al. [7] suggested decomposing singularities into position singularities and orientation singularities by partitioning the 6×7 Jacobian matrix. This method only needs to determine every singularity condition of the submatrices including two 3×4 matrixes, one 3×3 matrix, and one 3×3 zero matrix to obtain the singularity conditions of the robot. It can greatly reduce the computation complexity, but it can only be applied to the singularity analysis of the 7-DOF robot with the structure of the last three axes perpendicular to each other.

For a redundant robot with the structure of three consecutive parallel axes, this paper presents a method for singularity analysis based on matrix partitioning. Unchanging robot singularity, the analytical form of Jacobian matrix can be simplified by choosing an appropriate reference system and a reference point of the end-effector. Thereafter, the Jacobian matrix is reconstructed by matrix transformation after analyzing the characteristics of the simplified Jacobian matrix. On this basis, the Jacobian matrix is partitioned into four submatrices whose degradation

conditions will be discussed, and according to these degradation conditions, the singularity conditions and singular directions of the redundant robot are obtained. The approach has the following traits: (1) The Jacobian matrix is greatly simplified by choosing an appropriate reference system and a reference point of the end-effector; (2) the reason why there exists a 3×3 zero matrix in the simplified Jacobian matrix is analyzed in details, and it is suitable for any robot with the structure of three consecutive parallel axes; (3) the Jacobian matrix is reconstructed in a creative way according to the zero matrix, which provides the foundation for matrix partitioning; (4) the easy way of obtaining singularity conditions and singular directions, and the partition of Jacobian matrix provide convenience for singularity avoidance path planning.

This paper is organized as follows. Section 1 gives a brief overview of the research on the singularity analysis of redundant robots. Section 2 shows the studied robot configuration and simplifies the analytical form of the Jacobian matrix by choosing an appropriate reference system and a reference point of the end-effector. The singularity conditions and singular directions of a redundant robot with the structure of three consecutive parallel axes are obtained in Sect. 3. Section 4 verifies the correctness of the obtained singularity conditions and the singular direction corresponding to each singularity condition. Section 5 is the summary and conclusion of the work.

2 Simplification of the Jacobian Matrix

2.1 Robot Configuration

The robot studied in the paper is a 7-DOF robot with the structure of three consecutive parallel axes. Configuration like this has many advantages, such as operating flexibly, having a large span. Without loss of generality, the paper chooses the robot with the structure of axis 3, axis 4, axis 5 parallel to each other. The DH coordinate system is shown in Fig. 1, and DH parameters are given in Table 1.

Fig. 1 DH coordinate system of the studied robot

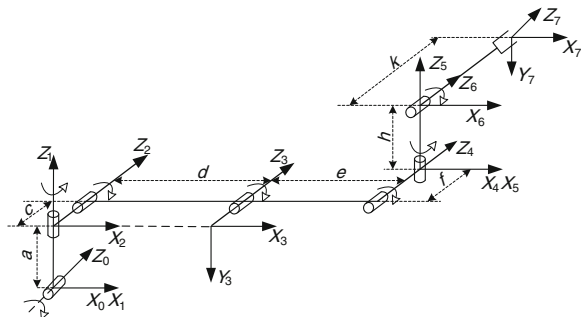


Table 1 DH parameters of the studied robot

link i	θ_i	d_i (mm)	a_{i-1} (mm)	α_{i-1}
1	$\theta_1(0)$	0	0	90°
2	$\theta_2(0)$	a	0	-90°
3	$\theta_3(0)$	0	d	0
4	$\theta_4(0)$	$c + f$	e	0
5	$\theta_5(0)$	0	0	90°
6	$\theta_6(0)$	h	0	-90°
7	$\theta_7(0)$	k	0	0

The Jacobian matrix $J_e^0 \in \mathbf{R}^{6 \times 7}$ of the manipulator relates the joint velocities in the joint space to the velocity of the end-effector in the Cartesian space (the subscript “ e ” denotes the end point, and the superscript “0” denotes vectors expressed in inertial coordinate system in subsequent description),

$$\dot{x}_e^0 = J_e^0 \cdot \dot{\theta} \tag{1}$$

where $\dot{x}_e^0 \in \mathbf{R}^6$ is the velocity of the end-effector and $\dot{\theta} \in \mathbf{R}^7$ is the joint angular velocity. Jacobian matrix J_e^0 can be obtained by vector product method [1]:

$$J_e^0 = \begin{bmatrix} z_i \times {}^n P_i \\ z_i \end{bmatrix}. \tag{2}$$

where z_i represents Z-axis direction vector of joint coordinate system i , ${}^n P_i$ is a vector from the origin of joint coordinate system i to the origin of the end coordinate system. They are both expressed in inertial coordinate system.

Since the Jacobian matrix of the studied robot is very complex in inertial coordinate system, it becomes very tough to analyze the singularity characteristics using matrix J_e^0 directly. Therefore, it is necessary to simplify the analytical form of the Jacobian matrix before singularity analysis. According to Eq. (2), the simplification of Jacobian matrix can be carried out in two ways: the simplification of z_i and the simplification of ${}^n P_i$. Vector z_i can be simplified by choosing an appropriate reference system, and ${}^n P_i$ can be simplified by choosing a proper reference point of the end-effector.

2.2 The Reference System

Usually, the base coordinate system is regarded as the robotic kinematics reference system. However, in order to simplify Jacobian matrix, the reference system should be chosen more reasonably. The principle of choosing the reference system is as follows: Select the intersection point formed by multiple axes as the origin of the reference system. Meanwhile, the Z-axis of the reference system should parallel to the joint axes as much as possible. According to the principle, there will be

more zero elements in Jacobian matrix, so the Jacobian matrix can be greatly simplified (9).

For the robot studied in this paper, the fifth joint coordinate system Σ_4 in Fig. 1 is chosen as the reference system Σ_{ref} . And then, the relationship between the Jacobian matrix J_e^{ref} and J_e^0 is shown as follows (the subscript “ref” denotes the reference point of the end-effector, and the superscript “ref” denotes vectors expressed in reference system Σ_{ref} in subsequent description):

$$J_e^0 = J_{\text{ref}}^0 \cdot J_e^{\text{ref}}. \quad (3)$$

where $J_{\text{ref}}^0 = \begin{bmatrix} {}^0R_{\text{ref}} & O_{3 \times 3} \\ O_{3 \times 3} & {}^0R_{\text{ref}} \end{bmatrix}$, ${}^0R_{\text{ref}}$ is the rotation matrix from Σ_{ref} to Σ_0 .

Since the determinant value of matrix J_{ref}^0 is 1, it can be found that the singularity of matrix J_e^0 is equivalent to that of matrix J_e^{ref} according to Eq. (3).

2.3 The Reference Point of the End-effector

In order to simplify the vector ${}^n P_i$ in Eq. (2), assume that the end-effector of the robot is enlarged infinitely, and select the intersection formed by the enlarged end-effector and the origin of coordinate system Σ_4 as the reference point of the end-effector, which is attached to the end-effector. It can be found that the actual end point vector x_e^{ref} and the end reference point vector $x_{e_ref}^{\text{ref}}$ have the following relationship:

$$x_e^{\text{ref}} = x_{\text{ref}}^{\text{ref}} + x_{e_ref}^{\text{ref}}. \quad (4)$$

where $x_{e_ref}^{\text{ref}}$ is the vector from the end reference point to the actual end point.

On this basis, the relationship between the end reference point velocity $v_{\text{ref}}^{\text{ref}}$ and actual end point velocity v_e^{ref} can be expressed as:

$$v_e^{\text{ref}} = v_{\text{ref}}^{\text{ref}} + \omega_{\text{ref}}^{\text{ref}} \times x_{e_ref}^{\text{ref}} = v_{\text{ref}}^{\text{ref}} - \hat{x}_{e_ref}^{\text{ref}} \cdot \omega_{\text{ref}}^{\text{ref}}. \quad (5)$$

where $\hat{x}_{e_ref}^{\text{ref}}$ is the antisymmetric matrix of $x_{e_ref}^{\text{ref}}$, and $\omega_{\text{ref}}^{\text{ref}}$ is the angular velocity of the end reference point.

Since the end reference point and the actual end point attach to the same rigid body, their angular velocities are the same.

$$\omega_{\text{ref}}^{\text{ref}} = \omega_e^{\text{ref}}. \quad (6)$$

where ω_e^{ref} is the angular velocity of the actual end point.

Combining Eq. (5) with Eq. (6), the actual end point velocity expressed in reference system is:

$$\dot{x}_e^{\text{ref}} = \begin{bmatrix} v_e^{\text{ref}} \\ \omega_e^{\text{ref}} \end{bmatrix} = \begin{bmatrix} I & -\hat{x}_{e_ref}^{\text{ref}} \\ O & I \end{bmatrix} \begin{bmatrix} v_{\text{ref}}^{\text{ref}} \\ \omega_{\text{ref}}^{\text{ref}} \end{bmatrix} = J_{e_ref}^{\text{ref}} \cdot \dot{x}_{\text{ref}}^{\text{ref}}. \quad (7)$$

And then combine Eq. (3) with Eq. (7), the relationship between the original Jacobian matrix J_e^0 and the simplified Jacobian matrix $J_{\text{ref}}^{\text{ref}}$ can be represented as:

$$J_e^0 = J_{\text{ref}}^0 \cdot J_e^{\text{ref}} = J_{\text{ref}}^0 \cdot J_{e_{\text{ref}}}^{\text{ref}} \cdot J_{\text{ref}}^{\text{ref}} \quad (8)$$

Because both the determinant values of J_{ref}^0 and $J_{e_{\text{ref}}}^{\text{ref}}$ are 1, the singularity of the original Jacobian matrix is equivalent to that of the simplified matrix.

For the 7-DOF robot with the structure of axis 3, axis 4, axis 5 parallel to each other, the Jacobian matrix can be expressed as (9) by the mentioned method.

$$J_{\text{ref}}^{\text{ref}} = \begin{bmatrix} C_2(aC_{345} + eS_5 + dS_{45}) - (c+f)S_2S_{345} & -(c+f)C_{345} & eS_5 + dS_{45} & eS_5 & 0 & 0 & 0 & -hC_6 \\ -(c+f)S_2C_{345} + C_2(-aS_{345} + eC_5 + dC_{45}) & (c+f)S_{345} & eC_5 + dC_{45} & eC_5 & 0 & 0 & 0 & 0 \\ -S_2(a - dS_3 - eS_{34}) & dC_3 + eC_{34} & 0 & 0 & 0 & 0 & 0 & -hS_6 \\ C_{345}S_2 & -S_{345} & 0 & 0 & 0 & 0 & 0 & -S_6 \\ -S_2S_{345} & -C_{345} & 0 & 0 & 0 & 0 & -1 & 0 \\ C_2 & 0 & 1 & 1 & 1 & 0 & 0 & C_6 \end{bmatrix} \quad (9)$$

In the above, $C_i, S_i, C_{ij}, S_{ij}, C_{ijk}, S_{ijk}$ represent $\text{Cos}(\theta_i), \text{Sin}(\theta_i), \text{Cos}(\theta_i + \theta_j), \text{Sin}(\theta_i + \theta_j), \text{Cos}(\theta_i + \theta_j + \theta_k)$ and $\text{Sin}(\theta_i + \theta_j + \theta_k)$, respectively.

Comparing the analytical form of original Jacobian matrix with the simplified Jacobian matrix, it is shown that the Jacobian matrix can be greatly simplified by choosing an appropriate reference system and a reference point of the end-effector. What's more, the simplification will provide the foundation for singularity analysis.

3 Singularity Analysis

On the basis of the Jacobian matrix simplification, it can be found that a 3×3 zero matrix exists in matrix $J_{\text{ref}}^{\text{ref}}$ according to Eq. (9). The zero matrix is resulted by the structure with three consecutive parallel axes.

3.1 The Characteristics of the Structure with Three Consecutive Parallel Axes

For a robot with the structure of three consecutive parallel axes, orientations of the joint coordinate axes corresponding to three consecutive parallel axes are usually consistent. Without any loss of generality, take the structure with axis 3, axis 4, axis 5 parallel to each other as an example, the directions of the coordinate axis Z_2, Z_3 and Z_4 are the same. Since the reference coordinate system Σ_{ref} and the joint coordinate system Σ_4 coincide with each other, the unit vectors of axis Z_2, Z_3 and Z_4 can be represented as $\{0, 0, 1\}^T$ with respect to Σ_{ref} . The Jacobian submatrix of

angular velocity corresponding to the structure with three consecutive parallel axes can be obtained based on Eq. (2):

$$J_{345}^{\omega} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}. \quad (10)$$

When solving the Jacobian submatrix of linear velocity, ${}^n P_i$ needs to be replaced by the vector from the origin of joint coordinate system i to the end reference point. Assume ${}^4 P_2 = \{a_x, a_y, a_z\}^T$, ${}^4 P_3 = \{b_x, b_y, b_z\}^T$ and ${}^4 P_4 = \{0, 0, 0\}^T$ can be easily obtained due to Σ_4 coinciding with Σ_{ref} . Substitute expression $z_2^{\text{ref}} = z_3^{\text{ref}} = z_4^{\text{ref}} = \{0, 0, 1\}^T$ into Eq. (2), the Jacobian submatrix of linear velocity corresponding to the structure with three consecutive parallel axes can be obtained as follows:

$$J_{345}^v = \begin{bmatrix} -a_y & -b_y & 0 \\ a_x & b_x & 0 \\ 0 & 0 & 0 \end{bmatrix}. \quad (11)$$

Thus, by merging Eqs. (10) and (11), the Eq. (12) can be constructed.

$$J_{345} = \begin{bmatrix} -a_y & a_x & 0 & 0 & 0 & 1 \\ -b_y & b_x & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}^T. \quad (12)$$

where J_{345} is the Jacobian submatrix corresponding to the structure with three consecutive parallel axes.

Through the derivation above, it can safely draw the conclusion that the submatrix (12) exists in the simplified Jacobian matrix of any robot with the structure of three consecutive parallel axes.

3.2 The Transformation of Jacobian Matrix

The characteristics of the structure with the three consecutive parallel axes can be used to simplify the singularity analysis. By elementary transformation for the Jacobian matrix $J_{\text{ref}}^{\text{ref}}$, the result of the transformation is shown as follows:

$$J_{\text{ref}}^{\text{ref}} = \begin{bmatrix} C_2 & 0 & C_6 & 0 & 1 & 1 & 1 \\ C_2(aC_{345} + eS_5 + dS_{45}) - (c+f)S_2S_{345} & -(c+f)C_{345} & -hC_6 & 0 & eS_5 + dS_{45} & eS_5 & 0 \\ -(c+f)S_2C_{345} + C_2(-aS_{345} + eC_5 + dC_{45}) & (c+f)S_{345} & 0 & 0 & eC_5 + dC_{45} & eC_5 & 0 \\ -S_2(a - dS_3 - eS_{34}) & dC_3 + eC_{34} & -hS_6 & 0 & 0 & 0 & 0 \\ C_{345}S_2 & -S_{345} & -S_6 & 0 & 0 & 0 & 0 \\ -S_2S_{345} & -C_{345} & 0 & -1 & 0 & 0 & 0 \end{bmatrix} \quad (13)$$

Since the elementary transformation of matrix does not change its singularity, the singularity of $J_{\text{ref}}^{\text{ref}'}$ instead of $J_{\text{ref}}^{\text{ref}}$ will be analyzed in the following discussion. According to Eq. (13), the relationship between the transformed velocity vector of the end reference point and the transformed joint velocity vector can be described as:

$$\dot{x}_{\text{ref}}^{\text{ref}'} = J_{\text{ref}}^{\text{ref}'} \cdot \dot{\theta}'. \quad (14)$$

where $\dot{x}_{\text{ref}}^{\text{ref}'} = \left\{ \omega_z^{\text{ref}'}, v_x^{\text{ref}'}, v_y^{\text{ref}'}, v_z^{\text{ref}'}, \omega_x^{\text{ref}'}, \omega_y^{\text{ref}'} \right\}^T$ is the transformed velocity vector, and $\dot{\theta}' = \left\{ \dot{\theta}_1, \dot{\theta}_2, \dot{\theta}_7, \dot{\theta}_6, \dot{\theta}_3, \dot{\theta}_4, \dot{\theta}_5 \right\}^T$ is the transformed joint velocity vector.

3.3 Singularity Analysis

On the basis of the transformation, it is necessary to partition the Jacobian matrix $J_{\text{ref}}^{\text{ref}'}$ to facilitate singularity analysis. Considering Eq. (13) having a 3×3 zero matrix, the partitioned matrix can be written as:

$$J_{\text{ref}}^{\text{ref}'} = \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & O_{3 \times 3} \end{bmatrix}. \quad (15)$$

3.3.1 The Last Three Rows Degradation Condition of $J_{\text{ref}}^{\text{ref}'}$

Since the last three rows degradation is only related to matrix $J_{21} \in \mathbf{R}^{3 \times 4}$, its singularity condition should be analyzed firstly. The determinant expression of matrix $J_{21} \cdot J_{21}^T$ is very complicated, so the singularity condition of matrix J_{21} will be analyzed by using its four submatrixes. All the four submatrix determinant values are listed in the following.

$$\begin{cases} \text{Det}(J_{21(1,2,3)}) = S_2 S_6 (h + aC_{345} + eS_5 + dS_{45}) \\ \text{Det}(J_{21(1,2,4)}) = S_2 (eC_5 + dC_{45} - aS_{345}) \\ \text{Det}(J_{21(1,3,4)}) = -S_2 S_6 (a + hC_{345} - dS_3 - eS_{34}) \\ \text{Det}(J_{21(2,3,4)}) = S_6 (dC_3 + eC_{34} + hS_{345}) \end{cases}. \quad (16)$$

where $J_{21(i,j,k)}$ represents the matrix composed of column i, j, k of J_{21} .

If all the submatrix determinant values equal zero simultaneously, the singularity of J_{21} can be judged. Set them equal zero simultaneously, and the singularity conditions are shown in Eq. (17).

Cond1, *Cond2*, and *Cond3* can be easily achieved, while *Cond4* is a little harder to be obtained. Observing the multiplication factors except S_2 and S_6 in Eq. (16), the relationships in Eq. (18) can be found:

$$\begin{cases} \text{Cond1} : S_2 = 0 \&\& S_6 = 0 \\ \text{Cond2} : S_2 = 0 \&\& dC_3 + eC_{34} + hS_{345} = 0 \\ \text{Cond3} : S_6 = 0 \&\& eC_5 + dC_{45} - aS_{345} = 0 \\ \text{Cond4} : dC_3 + eC_{34} + hS_{345} = 0 \&\& a + hC_{345} - dS_3 - eS_{34} = 0 \end{cases} \quad (17)$$

$$\begin{cases} h + aC_{345} + eS_5 + dS_{45} = S_{345}(dC_3 + eC_{34} + hS_{345}) + C_{345}(a + hC_{345} - dS_3 - eS_{34}) \\ eC_5 + dC_{45} - aS_{345} = C_{345}(dC_3 + eC_{34} + hS_{345}) - S_{345}(a + hC_{345} - dS_3 - eS_{34}) \end{cases} \quad (18)$$

From Eq. (18), we can see that the linear combination of $h + aC_{345} + eS_5 + dS_{45} = 0$ and $eC_5 + dC_{45} - aS_{345} = 0$ is equivalent to that of $dC_3 + eC_{34} + hS_{345} = 0$ and $a + hC_{345} - dS_3 - eS_{34} = 0$. Thus, *Cond4* can be drawn.

Substitute the four singularity conditions above into J_{21} , the base solution vectors in each singularity condition can be solved. Then, the singular direction corresponding to each singularity condition is listed as:

$$\begin{cases} U_{s1}^{\text{ref}} = \{0, 0, 0, S_{345}, dC_3 + eC_{34}, 0\}^T \\ U_{s2}^{\text{ref}} = \{0, 0, 0, 1, -h, 0\}^T \\ U_{s3}^{\text{ref}} = \{0, 0, 0, S_{345}, dC_3 + eC_{34}, 0\}^T \\ U_{s4}^{\text{ref}} = \{0, 0, 0, 1, -h, 0\}^T \end{cases} \quad (19)$$

3.3.2 The First Three Rows Degradation Condition of $J_{\text{ref}}^{\text{ref}'}$

According to Eq. (13), it can be found that all the elements of the fourth column in the first three rows are zeros. Thereby, the first three rows degradation only has the business with the determinant values of $J_{11(1,2,3)}$ and J_{12} . Both the determinant values are as follows:

$$\begin{cases} \text{Det}(J_{11(1,2,3)}) = -(c+f)C_6((c+f)S_2 - C_2(dC_3 + eC_{34} + hS_{345})) \\ \text{Det}(J_{12}) = deS_4 \end{cases} \quad (20)$$

When the first three rows degradation happens, both the equations in Eq. (20) should be equal to zero simultaneously. However, the conditions above are only need not sufficient conditions, another condition is necessary to ensure the degradation. The condition is that the degradation of $J_{11(1,2,3)}$ and J_{12} happens in the same row. Hence, the degradation condition of $[J_{11(1,2,3)}|J_{12}]$ will be discussed further.

Products can be obtained by multiplying the third row and the second row of $[J_{11(1,2,3)}|J_{12}]$ with S_5 and $-C_5$, respectively, and then replace the third row of matrix $[J_{11(1,2,3)}|J_{12}]$ with the sum of the products. So, the additional degradation condition can be got as $-(c+f)C_{34} = 0 \&\& aC_2C_{34} + dC_2S_4 - (c+f)S_2S_{34} = 0$. Combining Eq. (20), the first three rows singularity condition of $J_{\text{ref}}^{\text{ref}'}$ is shown in Eq. (21), and its singularity direction is in Eq. (22).

$$\begin{aligned} \text{Cond5} : S_4 = 0 \&\& C_6 = 0 \&\& - (c + f)C_{34} = 0 \\ \&\& aC_2C_{34} + dC_2S_4 - (c + f)S_2S_{34} = 0 \end{aligned} \tag{21}$$

$$U_{s5} = \{0, 0, 0, 0, -C_5, S_5\}^T. \tag{22}$$

So far, through the analysis and derivation above, the singularity conditions of the robot with the structure of three consecutive parallel axes and the corresponding singular directions are obtained. On the basis of Jacobian simplification, considering its characteristics and partitioning the Jacobian matrix, the singularity analysis is greatly simplified.

It is worth noting: In the process of singularity analysis above, the Jacobian matrix is partitioned, and the singular problem of original Jacobian matrix which consists of six-row vectors is converted into that of two matrixes which consist of three-row vectors separately. According to matrix theory, it may narrow the set of singularity conditions. However, it can be seen from Eq. (15), and the end reference point velocity is decoupled by matrix partitioning. So, the robot kinematics equation described in Eq. (1) can be transformed as:

$$\begin{cases} \dot{\theta}'_{(5,6,7)} = (J_{12})^{-1} \left[\dot{x}'_{\text{ref}(1,2,3)} - J_{11}\dot{\theta}'_{(1,2,3,4)} \right] \\ \dot{\theta}'_{(1,2,3,4)} = (J_{21})^\dagger \dot{x}'_{\text{ref}(4,5,6)} \end{cases} \tag{23}$$

where $\dot{\theta}'_{(i,j,\dots)}$ contains i th, j th, \dots elements of $\dot{\theta}'$, and $\dot{x}'_{\text{ref}((i,j,\dots))}$ consists of i th, j th, \dots elements of \dot{x}'_{ref} .

According to Eq. (23), robot path planning can be carried out only based on matrix J_{12} and J_{21} , which can greatly reduce the complexity.

4 Examples and Simulation Results

In order to verify the correctness of the singularity analysis method proposed in this paper, the following verifications are carried out. Assign the robot link parameters in Fig. 1: $a = c = f = h = k = 0.5$ m and $d = e = 2.5$ m. List five groups of robot joint angles satisfying the singularity conditions of (17) and (21), respectively, and solve the corresponding determinant values of $J_e^0 \cdot (J_e^0)^T$. And the relative results are shown in Table 2.

From Table 2, it can be found that all the determinant values of $J_e^0 \cdot (J_e^0)^T$ corresponding to the given joint angles approach to zero. Thus, it can be judged that robot singularity happens.

In the process of singularity analysis, because the simplified Jacobian matrix shows the relationship between the velocity of the end reference point and joint angular velocities in reference system, the singular direction describes the degradation direction of the end reference point with respect to reference system. Take

Table 2 Summary of examples

No	Singularity condition	Joint angle	Determinant value of $J_e^0 \cdot (J_e^0)^T$
Cond1	$S_2 = 0$	$\{25^\circ, 0^\circ, 34^\circ, 30^\circ, 45^\circ, 0^\circ, 0^\circ\}$	-8.23299×10^{-14}
Cond2	$S_2 = 0$	$\{25^\circ, 0^\circ, 20^\circ, 140^\circ, 20^\circ, 45^\circ, 0^\circ\}$	2.96681×10^{-30}
Cond3	$S_6 = 0$	$\{25^\circ, 10^\circ, 20^\circ, 140^\circ, 20^\circ, 0^\circ, 0^\circ\}$	3.20195×10^{-16}
Cond4	$dC_3 + eC_{34} + hS_{345} = 0$ $a + hC_{345} - dS_3 - eS_{34} = 0$	$\{25^\circ, 10^\circ, 0^\circ, 180^\circ, 0^\circ, 45^\circ, 0^\circ\}$	-1.40413×10^{-16}
Cond5	$S_4 = 0$ $aC_2C_{34} + dC_2S_4 - (c+f)S_2S_{34} = 0$	$\{25^\circ, 0^\circ, 90^\circ, 0^\circ, 45^\circ, 90^\circ, 0^\circ\}$	-1.0201×10^{-14}

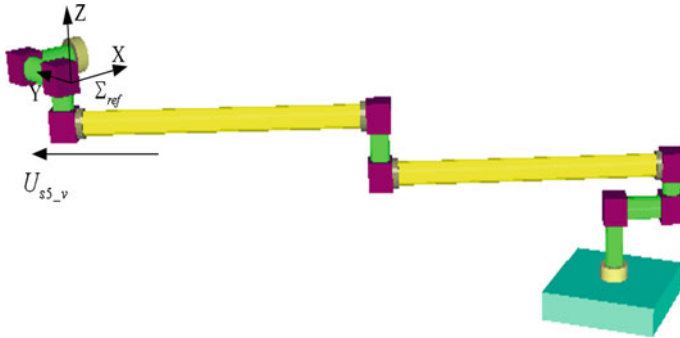


Fig. 2 Singular configuration corresponding to *Cond5*

Cond5 as an example, its singular direction which is expressed in Eq. (22) is shown in Fig. 2.

The simulation results verify the correctness and effectiveness of the proposed singularity analysis method for redundant robots with the structure of three consecutive parallel axes.

5 Summary and Conclusions

1. For redundant robots with the structure of three consecutive parallel axes, on the basis of Jacobian matrix simplification, a method based on matrix partitioning to analyze the robot singularity is proposed. All the singularity conditions and singular directions can be obtained by the method. The singularity analysis method can be used in control system of robots, providing the foundation for robot singularity avoidance path planning.
2. Unchanging robot singularity, the analytical form of the Jacobian matrix can be greatly simplified by choosing an appropriate reference system and a reference point of the end-effector, which provides the feasibility to carry out singularity analysis.
3. The simplified Jacobian matrix of redundant robot with the structure of three consecutive parallel axes contains a 3×3 zero matrix. Through matrix transformation, the velocity vector of the end reference point can be decoupled to simplify the singularity analysis.
4. The method proposed in this paper is not only applied to redundant robots with the structure of three consecutive parallel axes, but also applied to non-redundant robots with the same structure.

Acknowledgments This project is supported by National Key Basic Research Program of China (2013CB733005), National Natural Science Foundation of China (Grant No. 61175080), and the Fundamental Research Funds for the Central Universities (Grant No. BUPT2011rc0026).

References

1. Whitney DE (1969) Resolved motion rate control of manipulators and human prostheses. *J IEEE Trans Man Mach Syst* 10(2):47–53
2. Podhorodeski RP, Nokleby SB, Wittchen JD (2000) Resolving velocity-degenerate configurations (singularities) of redundant manipulators. In: Design engineering technical conferences and computers and information in engineering conference, p 10. Baltimore, USA
3. Nokleby SB, Podhorodeski RP (2001) Reciprocity-based resolution of velocity degeneracies (singularities) for redundant manipulators. *J Mech Mach Theor* 36:397–409
4. Nokleby SB, Podhorodeski RP (2004) Identifying multi-DOF-loss velocity degeneracies in kinematically-redundant manipulators. *J Mech Mach Theory* 39:201–213
5. Nokleby SB (2007) Singularity analysis of the canadarm2. *J Mech Mach Theor* 42:442–454
6. Waldron KJ, Reidy J (1986) A study of a kinematically redundant manipulator structure. In: IEEE international conference on robotics and automation
7. Cheng FT, Chen JS, Kung FC (1998) Study and resolution of singularities for a 7-DOF redundant manipulator. *J IEEE Trans Ind Electron* 45(3):469–480

Statistical Analysis of Stock Index Return Rate Distribution in Shanghai and Shenzhen Stock Market

Guan Li

Abstract This paper analyzes the distribution characteristics of Shanghai and Shenzhen (Shen-Hu) stock index return rate using statistic method. The result shows that: Shen-Hu stock index return rate shows left deviation and fat-tail distribution. Shen-Hu stock index rate do not follow the normal distribution and has no self-correlation. ARCH models based on T distribution of fixed 10 freedom investigate the distribution characteristics of Shen-Hu stock index return rate. The results show that EGARCH model can better describe the characteristics of Shen-Hu stock index return rate. News-driven asymmetry and “leverage effect” exist in Shen-Hu stock index return rate. In addition, there exists bilateral spillover effect in Shen-Hu market.

Keywords Return rate · ARCH models · Spillover effect

1 Background

The stock market in china began in the early 1990s and is at rise development stage. Because the stock market in china is still developing, there are more complex influential factors than markets in developed countries. There are several reasons accounting for the fluctuations of stock return rate during certain period. Generally, the time series of daily revenue of the Shanghai and Shenzhen (Shen-Hu) stock has non-normality and non-independence characteristic. Compared with normal distribution, the time series has heavy-tail characteristic and fluctuation clustering characteristic, all of which can be described by ARCH model. The stock markets in a same region are always linked because of the geographical proximity,

G. Li (✉)

Accounting Center of Tsinghua University, Beijing 100084, China
e-mail: lguan@mail.tsinghua.edu.cn

close economic relations, and similar policies. Thus, common information can affect the return and fluctuation of stock market in same region. Since both the stock exchanges of Shanghai and Shenzhen in the same country, studying the correlation and interactive development between the two stock markets is very important to analyze the structure of stock, determine the trend of stock, and estimate the transition of risk.

There have been lots of works concerned about the return rate of stock market. From the methodological point of view, ARCH models are usually used. Ruan [1] use Granger Causation theory [2] and GARCH-M model [3] to analyze the Shen-Hu stock return rate and fluctuation in empirical way. They concluded that there is strong correlation between two stocks and exist positive risk premium as well as asymmetry “Spillover Effect” in the transition of fluctuation. Yang [4] believes that the fluctuation of the stock will inevitably lead to big fluctuation of stock return rate during certain amount of time and small fluctuation in other time period. Since the traditional economic ways require a same variance which cannot be satisfied, the use of traditional regression model for those time series will lead to a distorted result. However, those characteristics can be described by ARCH models. Some people propose other methods. Lu [5] use other methods, such as Jarque-bera, Lilliefors, Cramer-von Mises, Watson, and Anderson-Darling to solve this problem and get the same result of ARCH.

In this paper, we use ARCH model to analyze the reasons that affect the return rate of Shen-Hu stock.

2 Experimental Results

2.1 Dataset Selection

This paper select 492 closing price samples of Shanghai and Shenzhen stock exchange from March 2, 2010 to March 9, 2011 (exclude the holidays) [6]. The logarithm yield of Shanghai stock y_t is calculated as $y_t = \ln x_t - \ln x_{t-1}$ where x_t is the closing price of Shanghai stock index at t day. The logarithm yield of Shenzhen stock index z_t is calculated as $z_t = \ln m_t - \ln m_{t-1}$ where m_t is the closing price of Shenzhen stock index at t day.

2.2 Statistic Properties of y_t and z_t

2.2.1 Distribution of y_t and z_t

Software eviews 5.0 is used to analyze the return rate y_t of the Shanghai stock index on day 492. The result is shown in Fig. 1.

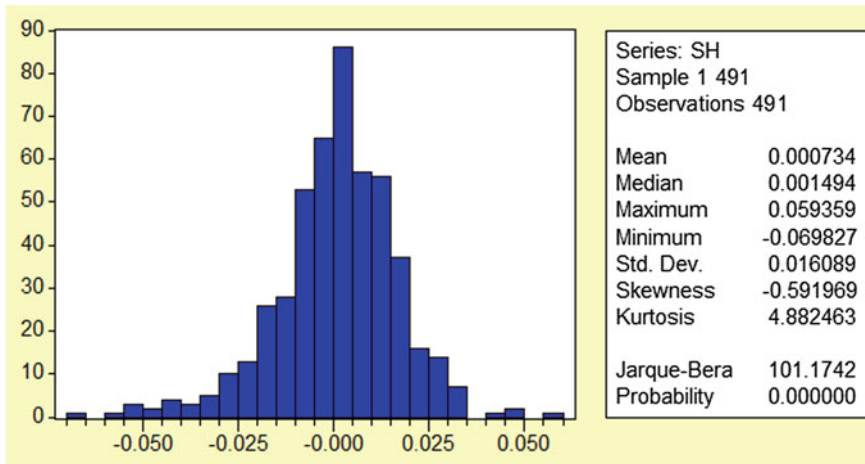


Fig. 1 Index return of Shanghai stock market

From Fig. 1, we can conclude that the mean of y_t is 0.000734, which is much lower compared with standard derivation 0.016089. Thus, the mean of y_t can be approximated as zero. The skewness $s = -0.591969 < 0$ and kurtosis $k = 4.882463 > 3$, where parameters of standard normal distribution is ($s = 0, k = 3$), shown as the left-skew and flat-tail distribution. The p value of JB test is zero. All of those show the series of return rate of Shanghai stock market y_t do not follow the normal distribution.

The return rate series z_t of the Shenzhen stock index is analyzed in a similar way and the result is shown in Fig. 2.

From Fig. 2, we can conclude that the mean of z_t is 0.001353, which is much lower compared with standard derivation 0.018319. Thus, the mean of z_t can be approximated as zero. The skewness $s = -0.728220 < 0$ and kurtosis $k = 4.554035 > 3$, where parameters of standard normal distribution are ($s = 0, k = 3$), shown as the left-skew and flat-tail distribution. The p value of JB test is zero. All of those show the series of return rate of Shenzhen stock market z_t do not follow the normal distribution. Since the standard derivation of Shenzhen stock return rate is larger than Shanghai, there is a larger fluctuation in Shenzhen stock market.

2.2.2 Correlation Test on Series Return Rate of y_t and z_t

If stochastic error is self-correlation, the estimation is not always correct. Thus, we need to verify whether the dataset has autocorrelation [7]. First, we do autoregressive analysis on the return rate series of Shanghai stock market, then, use Ljung-Box Q to fit the mean equation and calculate the residual error term. Finally,

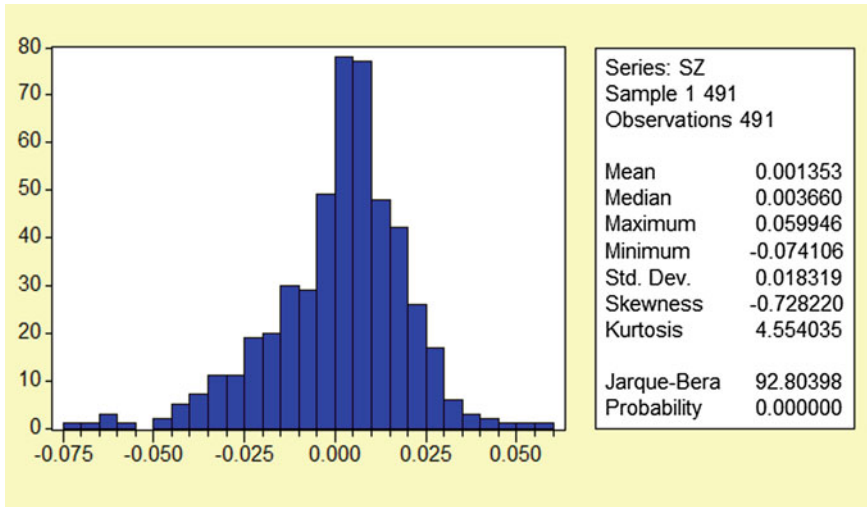


Fig. 2 Index return of Shenzhen stock market

<i>Autocorrelation</i>	<i>Partical Correlation</i>	<i>AC</i>	<i>PAC</i>	<i>Q-Stat</i>	<i>Prob</i>
1	1	0.025	0.025	0.3072	0.579
2	2	0.022	0.022	0.5478	0.760
3	3	0.057	0.056	2.1463	0.543
4	4	-	-	2.4745	0.649

Fig. 3 Autocorrelation test on residual error of Shanghai stock index

we test the autocorrelation of residual error term and select tenth-order lag. The result is shown in Fig. 3:

From Fig. 3, the P values of statistic Q are all above 0.05. Thus, there is no autocorrelation of the return rate of Shanghai stock market.

We can test the correlation of return rate series z_t of Shenzhen stock market in a similar way. The result is shown in Fig. 4.

From Fig. 4, the P values of statistic Q are all above 0.05. Thus, there is no autocorrelation of the return rate of Shenzhen stock market.







<i>Autocorrelation</i>	<i>Partical Correlation</i>	<i>AC</i>	<i>PAC</i>	<i>Q-Stat</i>	<i>Prob</i>	
		1	0.070	0.070	2.3640	0.124
		2	0.004	- 0.001	2.3720	0.305
		3	0.082	0.082	4.4985	0.133

Fig. 4 Autocorrelation test on residual error of Shenzhen stock index

2.2.3 Unit Root Test of Return Rate y_t and z_t

The purpose of unit root test on return rate series is to test whether the time series is stable. The stability of time series is defined as: the statistic law of time series will not change as time goes. In other words, the characteristic of random process generated from data will not change as time goes.

The unit root test of Shanghai return rate series is shown in Table 1:

From Table 1, we can see that the t value in ADF test is -21.80284 and p tends to zeros. When the series do not have lag period, intercept term or trend term, there is 99 % possibility to pass ADF test. Thus, the return rate series of Shanghai stock is also a stable time series.

The unit root test of Shenzhen return rate series z_t is shown in Table 2:

From Table 2, we can see that the t value in ADF test is -20.80326 and p tends to zeros. When the series do not have lag period, intercept term or trend term, there is 99 % possibility to pass ADF test. Thus, the return rate series z_t of Shenzhen stock is also a stable time series.

Table 1 Unit root test of Shanghai stock index

Augmented Dickey-Fuller test statistic	t-statistic	Prob.*
Test critical values: 1 % level	-21.80284	0
5 % level	-3.443469	
10 % level	-2.867219	
	-2.569857	

Table 2 Unit root test of Shenzhen stock index

Augmented Dickey-Fuller test statistic	t-statistic	Prob.*
Test critical values: 1 % level	-20.80326	0
5 % level	-3.443469	
10 % level	-0.867219	
	-2.569857	

Table 3 ARCH-LM test of residual error of Shanghai stock index

ARCH test			
F-statistic	7.676961	Probability	0.000051
Obs*R-squared	22.14905	Probability	0.000061

2.2.4 Whether the Return Rate Series y_t and z_t Have ARCH Effect

First, calculate a tenth-order lag regression of return rate of Shanghai stock market series, and then verify the residual error using ARCH-LM test and select the third-order lag. The result is shown in Table 3:

From Table 3, it shows that both the value of possibility p in F measure and value of possibility p in obs*R-squared is much lower than the significant level where $\alpha = 0.05$. Those indicate that the residual series of Shanghai return rate y_t has higher order ARCH effect.

The same way is applied to Shenzhen stock market. First, we calculate a tenth-order lag regression of return rate of Shenzhen stock market series, and then verify the residual error using ARCH-LM test [8] and select the second-order lag. The result is shown below.

From Table 4, it shows that both the value of possibility p in F measure and value of possibility p in obs*R-squared is much lower than the significant level where $\alpha = 0.05$. Those indicate that the residual series of Shenzhen return rate y_t has higher order ARCH effect.

2.3 Analyze ARCH Models

2.3.1 GARCH (1, 1) Model Based on T Distribution of Fixed 10 Freedom

The basic formula of this model is $\sigma_t^2 = c_0 + c_1 \mu_{t-1}^2 + c_2 \sigma_{t-1}^2$. In order to ensure the broad-balance of GARCH (1, 1) model, there is a constraint on parameters: $c_1 + c_2 < 1$. The GARCH (1, 1) model can describe lots of financial time series data. By using eviews 5.0, we can do model checking of GARCH (1, 1) on return rate series of Shanghai stock market y_t . The result is shown in Table 5:

From Table 5, the estimated equations of GARCH (1, 1) models for Shanghai are as follows:

Table 4 ARCH-LM test of residual error of Shenzhen stock index

ARCH test			
F-statistic	3.794012	Probability	0.023189
Obs*R-squared	7.516033	Probability	0.02333

Table 5 The GARCH (1,1) model checking of Shanghai stock index

	Coefficient	Std.error	z-statistic	Prob.
C	0.000714	0.00074	0.964485	0.3348
SH(-10)	0.055968	0.044657	1.253288	0.2101
<i>Variance Equation</i>				
C	1.26E - 05	7.27E - 06	1.730514	0.0835
RESID(-1)^2	0.043775	0.016821	2.602425	0.0093
GARCH(-1)	0.0903529	0.043019	21.0029	0
R-squared	0.004606	Mean dependent var		0.000691
Adjusted R-squared	-0.003759	S.D. dependent var		0.015891
S.E. of regression	0.015921	Akaike info criterion		-5.47615
Sum squared resid	0.120659	Schwarz criterion		-5.43274
Log likelihood	1322.015	F-statistic		0.55063
Durbin-Watson stat	1.945364	Prob(F-statistic)		0.698654

$$\sigma_t^2 = 1.26 \times 10^{-5} + 0.044 \mu_{t-1}^2 + 0.904 \sigma_{t-1}^2$$

$$Z = (1.731)(2.602)(21.003)$$

Log likelihood = 1322.015, AIC = -5.476152, SC = -5.432744

The same method can also be applied to the return rate z_t of Shenzhen stock market and the result is shown in Table 6:

From Table 6, the estimated equations of GARCH (1, 1) models for Shenzhen are as follows:

$$\sigma_t^2 = 2.86 \times 10^{-5} + 0.072 \mu_{t-1}^2 + 0.838 \sigma_{t-1}^2$$

$$Z = (1.835)(2.317)(11.521)$$

Log likelihood = 1259.084, AIC = -5.214, SC = -5.171

Table 6 The GARCH (1, 1) model checking of Shenzhen stock index

	Coefficient	Std.error	z-statistic	Prob.
C	0.001369	0.000909	1.505002	0.1323
SZ(-10)	0.090047	0.044511	2.023033	0.0431
<i>Variance Equation</i>				
C	2.86E - 05	1.56E - 05	1.835175	0.0665
RESID(-1)^2	0.072486	0.031286	2.31686	0.0205
GARCH(-1)	0.838019	0.072738	11.52107	0
R-squared	0.01041	Mean dependent var		0.001292
Adjusted R-squared	0.002094	S.D. dependent var		0.018176
S.E. of regression	0.018157	Akaike info criterion		-5.21449
Sum squared resid	0.15693	Schwarz criterion		-5.17108
Log likelihood	1259.084	F-statistic		1.251839
Durbin-Watson stat	1.85293	Prob(F-statistic)		0.288156

Table 7 The GARCH (1, 1)—M model checking of Shanghai stock index

	Coefficient	Std.error	z-statistic	Prob.
SQRT(GARCH)	0.175948	0.414494	0.42449	0.6712
C	-0.001925	0.00633	-0.304186	0.761
SH(-10)	0.056784	0.044927	1.263913	0.2063
<i>Variance equation</i>				
C	1.27E - 05	8.05E - 06	1.576751	0.1149
RESID(-1)^2	0.044098	0.018714	2.356384	0.0185
GARCH(-1)	0.903173	0.047788	18.89977	0
R-squared	0.004778	Mean dependent var		0.000691
Adjusted R-squared	-0.005698	S.D. dependent var		0.015891
S.E. of regression	0.015937	Akaike info criterion		-5.47139
Sum squared resid	0.120638	Schwarz criterion		-5.4193
Log likelihood	1321.869	F-statistic		0.456124
Durbin-Watson stat	1.939694	Prob(F-statistic)		0.808858

From estimated equations of GARCH (1, 1) on both Shanghai and Shenzhen stock markets, we can conclude that: the coefficients of Shanghai stock market in ARCH and GARCH is 0.044 and 0.904, respectively, the sum of which is near one. The coefficients of Shenzhen stock market in ARCH and GARCH is 0.072 and 0.803, respectively, the sum of which is near one. Both of those show the process of GARCH (1, 1) which is smooth. Thus, the fluctuations in both Shanghai and Shenzhen have a high persistence. In other words, once the stock return rate shows abnormal fluctuation because it is hard for impact to be eliminated in a short time.

2.3.2 GARCH (1, 1): M Model Based on T Distribution of Fixed 10 Freedom

The basic formula for GARCH (1, 1)—M is $\sigma_t^2 = c_0 + c_1 \mu_{t-1}^2 + c_2 \sigma_{t-1}^2$, which is same as GARCH (1, 1) as well as the parameters and constraints. By using eviews 5.0, and we can do model checking of GARCH (1, 1)—M on return rate series of Shanghai stock market y_t . The result is shown in Table 7:

From Table 7, the estimated equations of GARCH (1, 1)—M models for Shanghai are as follows:

$$\sigma_t^2 = 1.27 \times 10^{-5} + 0.044 \mu_{t-1}^2 + 0.903 \sigma_{t-1}^2$$

$$Z = (1.577)(2.356)(18.900)$$

$$\text{Log likelihood} = 1321.869, \quad \text{AIC} = -5.471, \quad \text{SC} = -5.419$$

The same method can also be applied to the return rate z_t of Shenzhen stock market and the result is shown in Table 8:

From Table 8, the estimated equations of GARCH (1, 1)—M models for Shenzhen are as follows:

Table 8 The GARCH (1, 1)—M model checking of Shenzhen stock index

	Coefficient	Std. error	z-Statistic	Prob.
SQRT(GARCH)	0.417868	0.372347	1.122252	0.2618
C	-0.005809	0.006475	-0.897239	0.3696
SZ(-10)	0.095413	0.044234	2.157028	0.031
<i>Variance Equation</i>				
C	2.93E - 05	1.80E - 05	1.62556	0.104
RESID(-1)^2	0.072704	0.034123	2.130623	0.0331
GARCH(-1)	0.835375	0.081961	10.19237	0
R-squared	0.013421	Mean dependent var		0.001292
Adjusted R-squared	0.003036	S.D. dependent var		0.018176
S.E. of regression	0.018149	Akaike info criterion		-5.21233
Sum squared resid	0.156452	Schwarz criterion		-5.16024
Log likelihood	1259.564	F-statistic		1.292327
Durbin-Watson stat	1.831732	Prob(F-statistic)		0.265875

$$\sigma_t^2 = 2.93 \times 10^{-5} + 0.073 \mu_{t-1}^2 + 0.835 \sigma_{t-1}^2$$

$$Z = (1.626)(2.131)(10.192)$$

$$\text{Log likelihood} = 1259.564, \quad \text{AIC} = -5.212, \quad \text{SC} = -5.160$$

From the above result, the SQRT coefficient ($\sqrt{\sigma_t^2}$), which represents risk degree, is 0.176 for Shanghai stock market and 0.418 for Shenzhen stock market respectively. The values of coefficient indicate that both stocks are significant. The risk and return rate are positive.

3 Conclusion

Based on rigorous statistical analysis, this paper selects various GARCH models [9, 10] to describe the distribution of Shen-Hu stock return series and draws the following conclusion: First, the return rate of Shen-Hu stock shows the left-skew and flat-tail distribution. Also the distribution do not follow a normal distribution and do not have self-correlation; Second, both the unit root test and ARCH effect test on return rate of Shen-Hu stock show the return rate series is stable. The residual series has a higher order ARCH effect; Third, both the sum of ARCH coefficient and GARCH coefficient on the Shen-Hu stock return rate is approximately one, which indicates the high volatility persistence in Shen-Hu stock market. In other words, once the stock return rate is fluctuating abnormally, the effect is hard to be eliminated; Fourth, from the test comparisons of various models on time series of Shen-Hu stock return rate, EGARCH model is better than TARCh model in analyzing the symmetry of news impact curve and determining the existence of “Leverage effect”; Fifth, the fluctuation of Shen-Hu stock exhibits

bilateral and symmetry spillover effect, indicating the two stock markets interact with each other.

Therefore, the finance methods under the assumption that the return rate series is independent and follows a normal distribution cannot be applied to the stock markets in China. Compared with foreign stock markets, Chinese stock market has its own characteristic, thus lots of methods cannot be directly used. In order to make a better analysis for Chinese stock market those methods need to be modified and reformed.

References

1. Ruan H, Xie J (2006) Instance study of returns and volatility in Shanghai and Shenzhen stock market. *Foreign Invest China* 2006(12):199 (in Chinese)
2. Chen S, Chen L, Liu Y (2003) An analysis of returns and volatility in Shanghai and Shenzhen stock market. *J Finance* 2003(7):80–85 (in Chinese)
3. Zhao F (2008) An empirical analysis on the earnings rate correlation between Shanghai and Shenzhen stock markets. *Sci Tech Inf Dev Econ* 2008(11):114–115 (in Chinese)
4. Yang H (2003) An empirical analysis of investment returns in Shenzhen stock market. *Stat Decis* 2003(1):67–68 (in Chinese)
5. Lu F (2004) Statistical character analysis of returns in Shanghai and Shenzhen stock market. *Stat Decis* 2004(12):13–14 (in Chinese)
6. Sun Q, Guang Z (2010) A comparative study of the return rate on stock market between China and America. *J Huaihua Univ* 2010(11):102–103 (in Chinese)
7. Ji X (2010) Renew on the theoretical literatures of foreign stock market bubbles. *J Shanxi Radio TV Univ* (7):78–80 (in Chinese)
8. Liu Q, Hao X (2008) Analysis on Shenzhen stock market returns characteristics. *Bus Cult* 2008(12):101–102 (in Chinese)
9. An Q, Guo X (2009) Study on stock return fluctuations based on GARCH-family models. *J Shandong Finance Inst* 2009(1):47–50 (in Chinese)
10. Zhong L, Zhang Q (2010) Correlation analysis on China stock market returns and economic net increase. *Bus China* 2010(6):79 (in Chinese)